NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS
CENTRE FOR FUNDAMENTAL SOCIOLOGY

RUSSIAN SOCIOLOGICAL REVIEW

✷

2024
Volume 23. Issue 4

ISSN 1728-1938 **Email:** puma7@yandex.ru **Web-site:** sociologica.hse.ru/en
Address: Staraya Basmannaya str., 21/4, Room A207, Moscow, Russian Federation 105066

About the Journal

The Russian Sociological Review is an academic peer-reviewed journal of theoretical, empirical and historical research in social sciences.

The Russian Sociological Review publishes four issues per year. Each issue includes original research papers, review articles and translations of contemporary and classical works in sociology, political theory and social philosophy.

Aims

- To provide a forum for fundamental issues of social sciences.
- To foster developments in social sciences by enriching theoretical language and vocabulary of social science and encourage a cross-disciplinary dialogue.
- To provide educational materials for the university-based scholars in order to advance teaching in social sciences.

Scope and Topics

The Russian Sociological Review invites scholars from all the social scientific disciplines to submit papers which address the fundamental issues of social sciences from various conceptual and methodological perspectives. Understood broadly the fundamental issues include but not limited to: social action and agency, social order, narrative, space and time, mobilities, power, etc.

The Russian Sociological Review covers the following domains of scholarship:

- Contemporary and classical social theory
- Theories of social order and social action
- Social methodology
- History of sociology
- Russian social theory
- Sociology of space
- Sociology of mobilities
- Social interaction
- Frame analysis
- Ethnomethodology and conversation analysis
- Cultural sociology
- Political sociology, philosophy and theory
- Narrative theory and analysis
- Human geography and urban studies

Our Audience

The Russian Sociological Review aims at both academic and non-academic audiences interested in the fundamental issues of social sciences. Its readership includes both junior and established scholars.

Subscription

The Russian Sociological Review is an open access electronic journal and is available online for free via http://sociologica.hse.ru/en.

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»
ЦЕНТР ФУНДАМЕНТАЛЬНОЙ СОЦИОЛОГИИ

СОЦИОЛОГИЧЕСКОЕ ОБОЗРЕНИЕ

✷

2024
Том 23. №4

ISSN 1728-1938 **Эл. почта:** puma7@yandex.ru **Веб-сайт:** sociologica.hse.ru
Адрес редакции: ул. Старая Басманная, д. 21/4, к. А207, Москва 105066

О журнале

«Социологическое обозрение» — академический рецензируемый журнал теоретических, эмпирических и исторических исследований в социальных науках. Журнал выходит четыре раза в год. В каждом выпуске публикуются оригинальные исследовательские статьи, обзоры и рефераты, переводы современных и классических работ в социологии, политической теории и социальной философии.

Цели

- Поддерживать дискуссии по фундаментальным проблемам социальных наук.
- Способствовать развитию и обогащению теоретического словаря и языка социальных наук через междисциплинарный диалог.
- Формировать корпус образовательных материалов для развития преподавания социальных наук.

Область исследований

Журнал «Социологическое обозрение» приглашает социальных исследователей присылать статьи, в которых рассматриваются фундаментальные проблемы социальных наук с различных концептуальных и методологических перспектив. Нас интересуют статьи, затрагивающие такие проблемы как социальное действие, социальный порядок, время и пространство, мобильности, власть, нарративы, события и т. д.

В частности, журнал «Социологическое обозрение» публикует статьи по следующим темам:

- Современные и классические социальные теории
- Теории социального порядка и социального действия
- Методология социального исследования
- История социологии
- Русская социальная мысль
- Социология пространства
- Социология мобильности
- Социальное взаимодействие
- Фрейм-анализ
- Этнометодология и конверсационный анализ
- Культурсоциология
- Политическая социология, философия и теория
- Нарративная теория и анализ
- Гуманитарная география и урбанистика

Наша аудитория

Журнал ориентирован на академическую и неакадемическую аудитории, заинтересованные в обсуждении фундаментальных проблем современного общества и социальных наук. Кроме того, публикуемые материалы (в частности, обзоры, рефераты и переводы) будут интересны студентам, преподавателям курсов по социальным наукам и другим ученым, участвующим в образовательном процессе.

Подписка

Журнал является электронным и распространяется бесплатно. Все статьи публикуются в открытом доступе на сайте: http://sociologica.hse.ru/. Чтобы получать сообщения о свежих выпусках, подпишитесь на рассылку журнала по адресу: farkhatdinov@gmail.com.

Contents

Содержание

РЕЦЕНЗИИ

A Socio-Cultural Perspectives on AI and the Global South

Guest Editor's Introduction

Arindam Das
Professor, School of Liberal Arts, Alliance University
Address: Chikkahagade Cross, Chandapura-Anekal Main Road, Anekal,
Bangalore — 562106
Karnataka, India
E-mail: arindam.das@alliance.edu.in

Glenn Muschert
Professor, Department of Public Health and Epidemiology
College of Medicine and Health Sciences, Khalifa University of Science and Technology
Address: PO Box 127788, Abu Dhabi, UAE
E-mail: glenn.muschert@ku.ac.ae

Introduction

In recent times, the ethical/morality concerns of AI (Arkin, 2009; Bello, Bringsjord, 2013) are no longer limited to its technical-industrial operationalities but even permeates the broader concerns (read imaginations) of politics and policy making (Brayne, 2017; Coeckelbergh, 2022; Joshi, 2024), society (Radhakrishnan, 2021; Gill, 2023; Pflanzer et al., 2023), economics (Vyshnevskyi et al., 2019; Qin et al., 2023; Capraro et al., 2024), and culture (Barron, 2023; Foka, Griffin, 2024; Barnes et al., 2024). The intersection of artificial intelligence (AI) and socio-cultural-economic contexts is a dynamic and evolving field of study (Feher, Katona, 2021). As AI technologies continue to advance, it is crucial to critically examine their impact on diverse societies, particularly those in the Global South — a segment of society that is not only economically and structurally less privileged but even socio-culturally less represented, racially discriminated, and historically subjugated and underdeveloped. AI development is highly concentrated in the techno-progressive Global North, and the research around AI and its associated social impact has primarily focused on more resource-unhindered Euro-American communities. Agencies, including the *World Economic Forum* (Yu et al., 2023), have highlighted this "AI divide between the Global North and the Global South." Although most early AI research and implementation took place in the West, the Global South holds a high potential to gain a great advantage from the technology. However, it is important to acknowledge that these nations in the Global South region face unique challenges in developing and applying AI. These challenges are related to digital literacy, internet penetration, basic electrical connections, and low infrastructure availability.

Additionally, there are uncertainties about the potential drawbacks of AI, potential biases and opacities in its applications, and ethical issues surrounding it. The Global South's foray into adopting and leveraging AI for agriculture, healthcare, education, climate action, poverty alleviation, and a general GDP gain faces challenges from data infrastructure/ecosystem, data governance, AI-adoptability, AI-usability, and AI-accessibility. At the same time, interaction between AI algorithms and the Global South communities demands investigation as to how cultural perspectives shape the ethical guidelines, how cultural biases get perpetuated or exacerbated due to AI advancements and how to mitigate them, how the new posthuman-ness impacts the socio-cultural matrix of the Global South, how AI in the Global South is reshaping and influencing the cultural production and economy, how AI is perpetuating neoliberal colonialism and capitalism in the Global South, and how the same is being negotiated/resisted by the subjects.

This Focus Issue acknowledges the transformative potential of AI, which has significantly impacted industry, politics, governance, economy, social interaction, and cultural dynamics in the Global South. However, many of these exuberant discourses of achievement and prognostications of potential success do not go without cautious neglect of the Global South's unique socio-cultural challenges and implications. This Focus Issue explores the multifaceted and complex relationship between AI and the socio-cultural-economic issues in the Global South, examining areas such as economic disparity, cultural hegemony, data logocentricity, and ethical governance. This Special Issue recognizes the need for intensive and sustainable research on the socio-cultural-politico-economic perspectives of AI in the Global South. This special issue sheds much-needed light on this topic: socio-cultural responses to AI adoption, AI inequalities, and how AI changes social-cultural life in the Global South region.

However, before we customarily introduce the articles and discussion in this collection, it would be much called for the readers to understand the significant critical themes/concerns that largely govern AI's interaction in the Global South's socio-economic-politico-cultural sphere.

AI and the Issues of Economic Disparities: Inequality Catalyst or Straddling Divide?

From a Global South perspective, the deployment and consecutive adoption of Global-North-ed AI have often caused an economic schism between the two socio-political halves of the globe. Indeed, AI has the capacity to polarise further the always-already existence of an unequal society (Lee, 2018; Dyer-Witheford et al., 2019; Crawford, 2021). Policy analysts and think tanks have often critiqued the Global North AI approaches for the Global South, predicting a scenario of greater economic divide that will exacerbate "global inequalities in the near term."

> The world of artificial intelligence (AI) is heavily lopsided. One American firm — Nvidia, the world's most valuable company — holds as much as 95% of the AI chip

> market. The $335 billion in private capital invested in American AI companies from 2013-2023 was three times more than in China, 11 times more than in the UK, and 30 times more than in India. And of the 109 most important machine learning models, 101 were made in the US, Western Europe, or China. Only two were made in a global South country (Egypt) (Laforge, 2024: n.pg.).

This geographic skew in AI production and governance (including computation and data storage and supply chain governance) and the high concentration of AI development in the Global North had created a higher demand for job reshoring, negatively impacting the "digital sectors in global South countries of foreign capital and income. There is evidence that this may already be occurring in Global South countries' IT sectors. Specifically, many of the areas of comparative advantage global South countries have developed in IT services are those that are highly exposed to AI-enabled automation. Many countries, such as India, have begun investing heavily in IT skills, but it is unclear if this will be sufficient to stem the potential outflow of capital and employment to the global North" (Jacobs, Tasin, 2024: n.p.g.). Indeed, the Global North's dominion of AI innovation (including high funding in research and development) through *big techs* like Amazon, Google, and Microsoft by extracting and leveraging data from the Global South ("without meaningful consent and fair compensation for the producers and sources of data" (Sadowski, 2019: 9)), and yet not extending equitable distribution benefits to the latter, had surely deepened the economic divide between the two. Here, one is reminded of *Facebook's* data extraction-appropriation forays through AI in Kenya with a total disregard for the African nation's data indigeneity and agency:

> A communiqué concerning Facebook, released on 1 September 2016, is a case in point, as it reflects, more broadly, the interests of US ICT companies in making business in Kenya and likely their willingness to contend the market to the Chinese influence. The document states that 'Zuckerberg said part of Facebook's overall strategy for Africa and Kenya is to understand what is happening on the continent and establish an entry point into the African economies for development. He said Facebook is committed to investing in connectivity.' This very general statement betrays the extent to which even a tech giant like Facebook seems to have long overlooked (for almost a decade) the potential of the whole SSA region, and Kenya's ICTs in particular. At the same time, it also shows Facebook's intention of making Kenya an entry point into the continent for its business, similar in manner to what Amazon recently announced (following the DPA approval) (Calzati, 2022: 279).

The extractive policy of data amassing from the Global South, whereby the latter is relegated to the role of agency-less suppliers rather than investors or beneficiaries, and commodifying the same to produce informational goods and services, can be termed AI colonialism/capitalism. Here, one is rightly reminded of the perspective extended by Engster and Moore about AI being a "specific capitalist mediation" (2020: 203, 212), contributing to the overarching agenda of capital accumulation. This lopsided dynamics between the Global North and the Global South sustain a cycle of reliance on AI, with the Global South merely serving as a repository for AI and its unavoidable consumption

rather than acting as an active creator of technology. Further, the neocolonial paradigm of various nation-states of the Global South is almost AI-washed in their "progressive" techno-modern agenda — unsuccessfully aping the Global North — to implement AI in labour-intensive sectors (viz., manufacturing and agriculture) only to further exacerbate unemployment and poverty.

Despite the discourses surrounding AI and its corresponding techno-capitalism, it is important to recognize the emerging yet powerful potential that the Global South's economic structure has been able to harness. The Global South has and is showing resilience in positively negotiating with AI. Leveraging AI for economic advancements, the Global South nations have utilized AI for more inclusive economic growth, addressing poverty and hunger. The Global South is all braced up to seize its moment of AI sovereignty from AI hegemony/colonialism. Remodelling global AI for local specificity needs and strategically aligning it with the developmental agendas of the nation and the Global South has shown reconcilable successes in particular cases. An ag-tech platform in Kenya, the "Hello Tractor," is revolutionizing the nation's agriscape for smallholder farmers. An AI-powered platform, it uses an IoT digital solution (dubbed "Uber for tractors") to connect tractor owners with farmers. This AI-enabled tractor-sharing service has been proven economically inclusive in fostering productivity by increasing the participation of marginalized and underserved farmers. Further, besides being financially inclusive, this AI-enabled model has proven sustainable and climate-smart (UNSGSA, 2023). Another case in point could be *Kuda* banking and AI-enabled fintech services in Nigeria that have helped millions of the country's "unbanked and underbanked" population. This AI-enabled banking platform has helped the underprivileged by giving access to financial rights and assisting the populace in economic activity. Kuda has helped reduce poverty by fostering financial inclusion (Empower Africa, 2023). Considering the progress of AI in the Global South, we understand that the agrarian economies are perhaps the most to benefit from harnessing AI power. Drone-equipped farming, optimized crop productivity, and intelligent energy consumption are all smart features that make agriculture more efficient (Wall et al., 2021). Despite these promises, the integration of AI is fraught with challenges in every single sector of the Global South. For instance, Wall et al. (2021) also warn about the ecological effects of AI usage, such as the potential harm to external wildlife, toxic emissions, and encroachment in external land due to rigid algorithms.

Socio-Cultural Implications of AI in the Global South: Imperialism and Beyond

The socio-cultural implications of AI in the Global South raise apprehensions about cultural imperialism. Within the historical technomodernity of Western imaginaries, the AI paradigm can be perceived to be "located within the colonial matrix of power," whether through its supply-chain model that reifies power imbalance between the West and its rest or "through an *international division of digital labor* that extracts value from the labour of workers in the majority world, generating profits for Western technology companies" (Muldoon, Wu, 2023: 3). Such imperialistic design of AI, informed by Western

value systems, in its production and supply process, fails to align with the cultural context of the Global South and is rather a hegemonic imposition of an episteme meant for acculturation and obeying in the labour market of the othered geography (Adas, 1989). Such ideological apparatus that produces knowledge/power and is deployed within the socio-technological sphere of the Global South leads to algorithmic governmentalities (Beer, 2016). The design and operation of AI systems are often informed by Western values, culture, languages, and norms that generally fail to align with the diverse cultural contexts of the Global South. For example, facial recognition systems that AI-identifies individuals based on the image of their facial features (gender, colour, age, race, emotion) have been disparaged for their racial and ethnic biases, performing poorly on darker-skinned individuals due to training and conditioning datasets predominantly composed of lighter-skinned individuals (Waelen, 2023). Such technological biases reinforce systemic discrimination and marginalization.

> Since recognition is considered to be constitutive of a person's ability to develop as an authentic and autonomous being and a condition for a just society, misrecognition should be seen as a threat to autonomy and a violation of justice. Alternatively, misrecognition could be perceived as a threat to well-being, since it hampers a person's flourishing. Hence, misrecognition is an ethical concern, because the psychological implications of misrecognition touch upon fundamental moral values and principles (Waelen, 2023: 218).

The production and deployment of *AI empires* through a Westernized system generates a computational power infrastructure that solidifies a Big-Tech structure of domination, essentially imperialistic in nature. A further extension of such imperialistic techno-hegemony and cultural domination is the integration of surveillance apparatus and unethical extortion of data. With the desire to sustain a cultural (Western) homogeneity through a digital panopticon, surveillance becomes systemic and integral to any AI-enabled industry (Das, Chanda, 2023: 192) where labour forces of the industry become "objects of information, never … subjects in communication" (Foucault, 1978: 108). An indigenous/tribal rights advocate and media professional, Nina Sangma (Garo), asserts the need for indigenizing emerging technologies and vents concern about AI surveillance in the lives of marginal indigenous communities: "One of the biggest concerns is the use of surveillance tech like Pegasus, which is being used to subvert democratic rights of citizens and free speech, including the targeting of journalists to curb freedom of the Press and citizens' right to information under the guise of national security. This, coupled with draconian laws like India's Armed Forces Special Powers Act, gives unbridled powers to the Army in so-called "disturbed areas" to maintain the status quo. These areas coincide with Indigenous lands where there is an Indigenous population, such as in Northeast India" (Sangma, 2024: n.pg.).

Furthermore, the widespread adoption of AI systems, promoted/justified globally based on the universal discourse of modernity, rationality, and objectivity, risks cultural homogenization. AI-driven content recommendation algorithms — such as those

employed by social media platforms — tend to prioritize popular or mainstream content, often originating from the Global North, at the expense of indigenous knowledge and local cultures. This erodes cultural diversity and undermines efforts to preserve languages, traditions, and practices unique to the Global South. As AI increasingly mediates communication and access to information, the preservation of cultural heritage becomes a critical concern. The creation of deterministic deepfakes by Generative AI may challenge the authenticity of indigenous cultural heritage, especially the intangible segment.

Nevertheless, the question remains if the future landscape of the Global South's computational power will depend on these transformative Westernized technologies' innovative diffusion (read cultural hegemony) to overcome disparities in equity and inclusivity. Also, whether the narrative of AI in the Global South has the potency to shift from dependency to agency remains. Notwithstanding the possibilities of systemic hegemony through AI that condition the socio-cultural ontology of subjects of the Global South, especially their indigenous and marginalized populations, there have been considerable explorations and negotiations of AI from the vantage point of the "othered" nations/communities/populace. Rethinking the otherwise imperialistic cultural dominion of the production-consumption continuum of Global North(ed) or Westernized AI, there are success stories of AI being indigenized and interpreted from a more local perspective and episteme. One such example is the creation of "TZ'IJK" — a "mestizo" (even postcolonial and hybrid) autonomous robotic agent inspired by Mayan creationist mythology — by Paula Gaetano Adi and Gustavo Crembil. An electronic art installation at a symposium at Simon Frazer University in 2015. According to the "Artist Statement," "TZ'IJK" is a strong response to the need for an alternative indigenized AI system beyond the Western cultural ideology-conditioned AI:

> Far from the utopias of smart, anthropomorphic and responsive machines, and inspired by the Maya's creationist mythology, TZ'IJK is a blind, deaf, and speechless autonomous robotic agent made from mud. Drawn from the lessons of mestizaje and motivated by Latin America's anthropophagic, cannibalistic, and hybrid nature, TZ'IJK proposes an alternative and disruptive approach to the development of embodied artificial life forms and advocates for the integration of high and low technological materials, processes, and cultures. Consisting of a large mud-covered sphere with an internal robotic mechanism, TZ'IJK establishes a non-reactive and unpredictable bodily interaction with the viewers. This creates the emergence of a new kind of synthetic agent that allows contradictions and ambiguity, complicating the traditional dichotomies of craft/technology, western/indigenous, modern/traditional, global/local, and developed/undeveloped (Adi, Crembil, 2015: n.pg.).

A more practical application that dehegemonizes AI can be the case of the Indian queer Adivasi engineer Aindriya Barua's (they/them) *ShhorAI*, "an AI-powered bot built to combat hate speech on social media, with a special focus on marginalised community safety" (Chakrapani, 2024: n.pg.).

Focus Issue Articles: Multifaceted Perspectives on AI and Global South

Bringing in the concerns of AI to the realm of urban studies, Borhan Sepehri et al. analyses how the urban infrastructure of Saudi Arabia may be AI-optimized to effectively become more inclusivist in implementing and enhancing the SDGs (with special emphasis on SDG5 [Gender Equality], SDG11 [Cultural Preservation and Heritage Protection], SDG4 [Skill-based Futuristic Education for Youth], SDG8 [Decent Job and Economic Growth for Youth], SDG6 [Access to Safe Water, Water Management, and Sanitary Governance], SDGs7/13/14/15 ["Reducing Climate Change, Creating Sustainable Solutions, Forecasting Solar Photovoltaic Power, Improving Renewable Energy Efficiency, Increasing Climate Flexibility, Weather Forecasting, Water Resource Management, and Promoting Agricultural Practices Resistant to Climate and Food Security"]). This the authors achieve through a narrative review method and their research finally emphasizes AI for Saudi Arabia that must unfailingly "prioritize social, religious, and cultural characteristics and values" "highly compatible with Saudi society". This research is indeed a recognition of the transformative potential for AI in the Global South and no less an assertion for the decolonization of the same to suit the local agendas.

In the following article by Demirel et al., we counter the power imbalance in the socio-cultural representations and perceptions of generative AI (ChatGPT) between the Global North and Global South, as captured on Twitter. The text analysis employed in the research indicates that the Global North focuses "more on sectoral applications and technical aspects, while the Global South evaluates ChatGPT within local language and cultural contexts. The findings demonstrate that socio-cultural differences and technological development levels between regions are reflected in the social representations of ChatGPT." Based on the findings, the research further sheds light on the Global South's concerns about AI-related privacy issues, cybersecurity, fake news, and consequent culturally compatible solutions. The findings further indicate the Global South's perception of Gen-AI as a threat to employment, and this is to be ascribed to the disadvantageous socio-cultural-economic context of Third World nations like Turkey and India (two representational countries in the research). This "AI anxiety" (Li, Huang, 2020) of the Global South, vis-à-vis the relatively progressive and futuristic association of the Global North with AI, could no less be perceived from an ontological dimension. The Global South realities and its complex existential AI anxieties across demographics are an agenda that future researchers may probe through this article.

Next, Sinha's article captures a more pronounced version of this AI anxiety. Sinha talks about the neocolonial surveillance model in India (through the nation-state's deploying of the system of *Aadhaar* — an AI-driven biometric identification initiative) that has led to data (neo)colonialism, which in turn conditions/normalizes through a disciplinary knowledge/power the individual and collective identities of the subjects of the nation-state. However, Sinha is not altogether pessimistic, for she projects possibilities of Bhabhaesque hybridity-resistance (Bhabha, 1984), whereby "Aadhaar could be reimagined as a tool for equitable governance."

Varghese and Rani's article continues with this trope of perceiving AI as a colonial construct and its functioning/mechanism as that of "digital orientalism." Through a systematic analysis of 270 AI-generated images, "this study investigates how contemporary artificial intelligence image generation systems interpret and reproduce Indian cultural elements." It concludes that such representations lack cultural sensitivity and are lopsided with an oriental bias. Visual stereotyping and algorithmic cultural reductionism of Indian society by the Western AI system perpetuates power imbalance whereby the latter almost engenders scopic violence and representational damage on the former.

Finally, Shomotova et al.'s article, "The Impact of Socio-Cultural and Demographic Factors on Gen AI Accessibility, Usability, and Applicability in the UAE," is a more practical concern about the feasibility of Westernized AI within a Global South scenario, given its multicultural/multiethnic composition and correspondingly nuanced issues of demography. The article's mixed-methods research design showed strong links between how Gen AI uses AI and demographic and educational factors. Since they appreciated real-time feedback and time-saving features of Gen AI tools, many students stated that accessibility was crucial. However, they ran into problems like prompt sensitivity and the need to verify output. Although the Gen AI Applicability results highlighted ChatGPT's assistance in content creation, language improvement, and academic material organization, they also noted that it struggles to adhere to assignment-specific guidelines. This study adds to the body of literature by examining the sociodemographic elements that affect the adoption of Gen AI in a culturally diverse environment such as the United Arab Emirates.

This Focus Issue culminates with a thought leader Roundtable, whereby the Editors of the Special Issue engage in "a critical dialogue between leading scholars in the fields of Sociology, Critical Communication Studies, Cultural Studies, Critical Management Studies, and Sustainability Studies to explore the challenges that Global South navigates in its adoption of AI." The conversation does not merely probe the concerns around power disparities in the AI paradigm of the Global South and the Global North. However, it even suggests decolonizing the AI system by prioritizing "human rights, ethics, equity, inclusivity, and resilience."

Conclusion

Thus, this Focus Issue either non-obliquely or tangentially projects an aspiration for decolonizing the AI of the Global North that is deeply entangled in the global dynamics of power, capital, and culture, perpetuating a cycle of neo-imperialism that mirrors historical patterns of colonial domination and reflects hegemonic structures that disproportionately benefit the Global North. By extracting data, exploiting labor, and imposing Western epistemologies, AI technologies have not only deepened global economic disparities but have also eroded cultural diversity and reinforced systemic biases against the Global South, especially their marginalized communities. Indeed, a strategic decolonial AI framework must prioritize the agency of the Global South in shaping technological

futures. This includes fostering equitable access to AI infrastructure, ensuring culturally sensitive design, and recognizing indigenous epistemologies as valid knowledge systems. Moreover, critical engagement with AI must interrogate its deployment's ethical, economic, and socio-political implications, aiming to dismantle the structures of power that perpetuate dependency and inequality. The future of AI in the Global South lies in crafting context-sensitive, culture-sensitive, inclusive, and equitable models of technological development — a deconstructed AI of its own. Decolonizing AI requires a fundamental shift — from a system that extracts and exploits to one that empowers and sustains by centering the Global South's voices, needs, and aspirations.

It has been our sincere honor to convene this diverse group of scholars and to play a role in curating its contributions. While AI remains a hot topic in scholarship and economic news, we note the inherent biases of the Global North and the West from its inception. Therein is the need and the justification for such investigations, as inequalities likely beget further inequalities, biases, and injustices. We are grateful to our contributors for entrusting us to shepherd their work through the review process. Similarly, we are grateful to peer reviewers whose selfless contributions remain anonymous despite playing a necessary and pivotal part in the process. Finally, we thank the Editors and Editorial Board of the *Russian Sociological Review* for advising us through this process and entrusting us with curating the content for this Special Issue.

Professor Arindam Das and Professor Glenn Muschert
December 2024

References

Adas M. (1989) *Machines as the Measure of Men: Science, Technology, and Ideologies of Western Dominance,* Ithaca and London: Cornell University Press.

Adi P. G., Crembil G. (2015) "Paula Gaetano Adi, Gustavo Crembil: TZ'IJK," *ISEA Symposium Archives*. Accessed on 25/12/2024. https://isea-archives.siggraph.org/art-events/paula-gaetano-adi-gustavo-crembil-tzijk/.

Arkin R. (2009) *Governing Lethal Behavior in Autonomous Robots*, London: Chapman and Hall/CRC Imprint, Taylor and Francis Group.

Barnes A. J., Zhang Y., Valenzuela A. (2024) AI and Culture: Culturally Dependent Responses to AI Systems. *Current Opinion in Psychology*, no 58, p.101838.

Barron L. (2023) *AI and Popular Culture,* Bingley: Emerald Publishing.

Beer D. (2016) The Social Power of Algorithms. *Information, Communication & Society*, vol. 20, no 1, pp. 1–13.

Bello P., Bringsjord S. (2013) On How to Build a Moral Machine. *Topoi*, vol. 32, no 2, pp. 251–266.

Bhabha H. K. (1994) *The Location of Culture,* London: Routledge.

Brayne S. (2017) Big Data Surveillance: The Case of Policing. *American Sociological Review*, vol. 82, no 5, pp. 977–1008.

Bringsjord S., Govindarajulu N. S. (2018) Artificial Intelligence. *Stanford Encyclopedia of Philosophy*. Accessed 26/12/2024. https://plato.stanford.edu/entries/artificial-intelligence/.

Calzati S. (2022) 'Data Sovereignty' or 'Data Colonialism'? Exploring the Chinese Involvement in Africa's ICTs: A Document Review on Kenya. *Journal of Contemporary African Studies*, vol. 40, no 2, pp. 270–285.

Capraro V., Lentsch A., Acemoglu D. et al. (2024) The Impact of Generative Artificial Intelligence on Socioeconomic Inequalities and Policy Making. *PNAS Nexus,* vol. 3, no 6, p. 191.

Chakrapani S. (2024) This Queer, Adivasi Engineer has a Next-gen Solution to Homophobic Trolls. *Social Story*. Accessed on 24/12/2024. https://yourstory.com/socialstory/2024/06/queer-adivasi-engineer-next-gen-solution-homophobic-trolls.

Coeckelbergh M. (2022) *The Political Philosophy of AI,* Cambridge: Polity Press.

Crawford K. (2021) *Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence*, New Haven: Yale University Press.

Das A., Chanda D. (2023) To Trust or Not to Trust Cybots: Ethical Dilemmas in the Posthuman Organization." *New Horizons for Industry 4.0 in Modern Business*. (eds. A. Nayyar, M. Naved, R. Rameshwar), Cham: Springer, pp. 189-208.

Dyer-Witheford N., Kjøsen A. M., Steinhoff J. (2019) *Inhuman Power. Artificial Intelligence and the Future of Capitalism*, London: Pluto Press.

Empower Africa (2023) Nigerian Fintech Startup Kuda Hits 6 Million Customer Milestone as it Continues to Redefine Financial Inclusion in Africa. 24/07/2023. Accessed on 26/12/2024. https://empowerafrica.com/nigerian-fintech-kuda-hits-6-million-customer-milestone-as-it-continues-to-redefine-financial-inclusion-in-africa/.

Engster F., Moore P. V. (2020) The Search for (Artificial) Intelligence in Capitalism. *Capital & Class*, no 44, pp. 201–218.

Feher K., Katona A. I. (2021) Fifteen Shadows of Socio-cultural AI: A Systematic Review and Future Perspectives. *Futures*, no 132, p. 102817.

Foka A., Griffin G. (2024) AI, Cultural Heritage, and Bias: Some Key Queries That Arise from the Use of GenAI. *Heritage*, vol.7, no 11, pp. 6125-6136.

Foucault M. (1978) *The History of Sexuality Vol. 1: An Introduction*. (R. Hurley, Trans.), London: Random House, Inc.

Gill K. S. (2023) Seeing beyond the Lens of Platonic Embodiment. *AI & Society*, no 38, pp. 1261–1266.

Jacobs J., Tasin F. (2024) How the Global South may Pay the Cost of AI Development. *OMFIF*. Accessed on 26/12/2024. https://www.omfif.org/2024/07/how-the-global-south-may-pay-the-cost-of-ai-development/.

Joshi D. (2024) AI governance in India — Law, Policy and Political Economy. *Communication Research and Practice*, vol. 10, no 3, pp. 328–339.

Laforge G. (2024) The Danger of Imposing Global North Approaches to AI Governance on the Global South. *TechPolicy, Press*. Accessed on 26/12/2024. https://www.techpolicy.press/the-dangers-of-imposing-global-north-approaches-to-ai-governance-on-the-global-south/.

Lee K. F. (2018) AI superpowers. China, Silicon Valley and the New World Order, Boston: Houghton Mifflin Harcourt.

Li J., Huang J. (2020) Dimensions of Artificial Intelligence Anxiety based on the Integrated Fear Acquisition Theory. *Technology in Society*, no 63, p. 101410.

Muldoon J., Wu B. A. (2023) Artificial Intelligence in the Colonial Matrix of Power. *Philosophy and Technology,* no 36, Article no 80. https://doi.org/10.1007/s13347-023-00687-8

Pflanzer M., Dubljević V., Bauer W. A. (2023) Embedding AI in Society: Ethics, Policy, Governance, and Impacts. *AI & Society*, no 38, pp. 1267–1271.

Qin Y., Xu Z., Wang X., et al. (2024) Artificial Intelligence and Economic Development: An Evolutionary Investigation and Systematic Review. *Journal of Knowledge Economy*, no 15, pp. 1736–1770.

Radhakrishnan R. (2021) Experiments with Social Good: Feminist Critiques of Artificial Intelligence in Healthcare in India. *Catalyst: Feminism, Theory, Technoscience*, vol. 7, no 2, Article 2.

Sadowski J. (2019) When Data is Capital: Datafication, Accumulation, and Extraction. *Big Data & Society*, vol. 6, no 1, pp. 1–12.

Sangma N. (2024) Artificial Intelligence and Indigenous Peoples' Realities. *Cultural Survival*. Accessed on 24/12/2024. https://www.culturalsurvival.org/publications/cultural-survival-quarterly/artificial-intelligence-and-indigenous-peoples-realities.

UNSGSA (United Nations Secretary General's Special Advocate for Financial Health) (2023) Hello Tractor is Revolutionizing Farming and Fueling Economic Empowerment Through Digital Financing for Smallholders in Kenya." 07/11/2023. Accessed on 26/12/2024. https://www.unsgsa.org/stories/hello-tractor-revolutionizing-farming-and-fueling-economic-empowerment-through-digital-financing-smallholders-kenya.

Vyshnevskyi O., Liashenko V., Amosha O. (2019) The Impact of Industry 4.0 and AI on Economic Growth," *Scientific Papers of Silesian University of Technology Organization and Management Series,* no 9, pp. 391–400.

Waelen R. A. (2023) The Struggle for Recognition in the Age of Facial Recognition Technology. *AI Ethics*, no 3, pp. 215–222.

Wall P. J., Saxena D., Brown S. (2021) Artificial Intelligence in the Global South (AI4D): Potential and Risk." *Proceedings of the 1st Virtual Conference on Implications of Information and Digital Technologies for Development, 2021*. In *arxiv.org*. Accessed on 24/12/2024. https://arxiv.org/pdf/2108.10093v1

Yu. D., Rosenfeld H., Gupta A. (2023) The 'AI Divide' between the Global North and the Global South. *World Economic Forum*. Accessed on 26/12/2024. https://www.weforum.org/stories/2023/01/davos23-ai-divide-global-north-global-south/

doi: 10.17323/1728-192x-2024-4-20-47

Artificial Intelligence Role in Promoting Saudi Arabia's Smart Cities: Addressing SDGs for Socio-Cultural Challenges

Borhan Sepehri
Department of Urban Planning & Design, Tarbiat Modares University,
Address: P. O.Box: 14115-111, Jalal Al Ahmad Highway Nasr
Tehran, Iran
E-mail: borhanspehri@gmail.com

Abdulaziz I. Almulhim
Department of Urban and Regional Planning, College of Architecture and Planning
Imam Abdulrahman Bin Faisal University
Address: P. O. Box 1982, Dammam, 31451, Saudi Arabia
E-mail: aialmulhim@iau.edu.sa

Mohammad Anvar Adibhesami
Department of Architecture, School of Architecture and Environmental Design
Iran University of Science and Technology
Address: Narmak, Tehran, Iran
E-mail: adibhesami@hotmail.com

Saeed Makaremi
Department of Urban Planning and Design, Shahid Beheshti University
Address: Shahid Shahriari Square, Evin Tehran, Iran
Postal Code: 1983969411
E-mail: s.makaremi@alumni.sbu.ac.ir

Farzaneh Ejazi
Department of Social Science, Faculty of Human Sciences
Shahed University
Address: Khalij Fars Expressway Tehran, Iran Postal Code: 3319118651
E-mail: farzanehejazi6@gmail.com

The study explores the use of artificial intelligence (AI) to address socio-cultural challenges in Saudi Arabia while promoting Sustainable Development Goals (SDGs). Using a structured narrative review method with a critical approach, the study highlights AI's versatility in analyzing complex social phenomena, understanding human behavior, and optimizing urban infrastructure. With its unique socio-cultural challenges, Saudi Arabia aims for sustainable development through Saudi Vision 2030 and various smart city projects, emphasizing the importance of addressing challenges like gender equality, cultural preservation, an increasing youth population, rapid urbanization, and climate change.

The study identifies ten AI applications and models to address these challenges and promote relevant SDGs across six areas: Predictive Analytics and Forecasting, Optimization and Decision Support, Natural Language Processing, Computer Vision, Generative AI, and Geospatial AI. These AI models can help address issues like gender equality, youth education, and employment, as well as optimize water management, energy use, and urban planning to address rapid urbanization and climate change challenges.

By aligning AI development with the goals outlined in the SDGs, Saudi Arabia can unlock the potential of AI to create sustainable, resilient, and inclusive smart cities that effectively address

socio-cultural challenges. However, the study emphasizes the necessity of customizing AI applications in smart cities based on Saudi Arabia's religious and cultural values to ensure ethical and culturally sensitive implementations. The findings of this study hold relevance not only for Saudi Arabia but also for other countries facing similar challenges. The study provides practical recommendations for policymakers, urban planners, and technology experts to leverage AI effectively for sustainable development. It also outlines future research directions to address the limitations identified, such as exploring implementation challenges and ethical considerations.

Keywords: AI; SDGs; Socio-cultural challenges; Smart Cities; Saudi Arabia; Saudi Vision 2030; NEOM

Introduction

Artificial Intelligence (AI) has significantly impacted social and urban sciences, analyzing social phenomena and optimizing urban infrastructure (Karimi et al., 2024). It is also instrumental in urban planning, resource allocation, and urban infrastructure optimization (Adibhesami et al., 2024). AI is widely acknowledged as a valuable tool for effectively implementing and enhancing sustainable development goals (SDGs) (Mashhood et al., 2023). Implementing SDGS can create jobs, increase income levels, provide innovative solutions, improve infrastructure productivity, provide quality services like education and healthcare, promote economic growth, generate knowledge centers, strengthen social and cultural vibration, and improve living standards (Dhanraj et al., 2023; Gómez-Villarino, Briz, 2022). AI has the potential to enhance sustainability in communities and cities while maximizing positive impact (Mashhood et al., 2023). Research indicates that AI can expedite progress toward the United Nations' SDGs, promoting sectors such as environmental monitoring, energy efficiency, climate change mitigation, and healthcare (M. Yadav, Singh, 2023). However, a careful, ethical, and systematic approach to AI technology development and integration is required to harness this potential (Chen et al., 2023).

"Smart cities" refer to urban areas that leverage digital technologies and data analytics to enhance the quality of life for residents, optimize resource management, and ensure sustainable development (Kondepudi, Kondepudi, 2018). AI encompasses a range of technologies that enable machines to perform tasks that typically require human intelligence, such as learning, reasoning, problem-solving, and decision-making. In this study, AI is presented as a tool that supports the achievement of SDGs by providing innovative solutions to socio-cultural challenges (Ziesche et al., 2023).

Therefore, AI can be introduced as a suitable tool to strengthen the SDGs in a particular society, which will lead to solving the socio-cultural problems of that society. Saudi Arabia is one of the countries that is rapidly modernizing and struggling with various socio-cultural challenges. By implementing the Saudi Arabia 2030 vision for sustainable development, Saudi Arabia has taken essential steps to create a sustainable society (Alnasser, Musallat, 2022; Balabel, Alwetaishi, 2021). In addition, various smart city projects such as NEOM and Visionary Downtown Riyadh increase the intelligence and sustainability of Saudi communities. NEOM is a groundbreaking urban development project in Saudi Arabia, aimed at creating a futuristic city that integrates advanced technolo-

gies and sustainable living. AI is used in NEOM for infrastructure management, cyber security, and education (Alismail, Faridi, 2021; DARRAJ, 2019). At the same time, Saudi Arabia is facing challenges such as rapid urbanization, a growing young population, and demand for sustainable growth, public facilities, housing, and infrastructure (Abubakar, Aina, 2018; B. S. Alotaibi et al., 2024). Furthermore, In Saudi Arabia, various case studies highlight the application of AI to address socio-cultural. AI is enhancing e-learning experiences, particularly for women, thereby improving access to quality education. In healthcare, AI optimizes dental care and medication management, increasing accessibility and safety. Additionally, AI is utilized in disaster-resilient architectural designs that respect cultural heritage and in forensic science to strengthen justice systems. These initiatives demonstrate how AI can effectively tackle socio-cultural issues while promoting sustainable development in the region (Almajed et al., 2024; Luppicini, Walabe, 2021).

However, the limited research on SDGs, socio-cultural challenges, AI, and sustainability in Saudi Arabia is insufficient. For example, research shows explicitly that SDGs 1 and 3 are particularly relevant for addressing social challenges in Saudi Arabia (Guermazi, Gharbi, 2024). Also, studies collectively emphasize the importance of AI in driving sustainable practices in various sectors of Saudi Arabia, which aligns with the country's strategic goals for economic diversification and environmental stewardship (N. S. Alotaibi, Alshehri, 2023; Mutambik, 2024). However, Saudi Arabia has a significant research gap regarding the role of AI in implementing SDGs and solving socio-cultural challenges. In this regard, a narrative review method with a critical approach can be useful because increasing understanding, improving decision-making, promoting responsibility, and advancing knowledge in various fields is vital, and these are possible with the help of a narrative review method.

In light of these developments, this study explores the utilization of AI to address socio-cultural challenges in Saudi Arabia by promoting SDGs. The following goals guide this study: Identify the most important socio-cultural challenges in Saudi Arabia, analyze which SDGs would help address these challenges, review AI applications that would help achieve these SDGs, and summarize AI applications that may contribute to overcoming current socio-cultural challenges in Saudi Arabia. By clearly defining these goals, the study ensures coherence among the research objectives, methodology, results, and conclusions. This study adopts a theoretical framework that posits AI as a facilitator in achieving SDGs, thereby mitigating socio-cultural challenges. This framework is grounded in the theory that technological advancements can drive significant socio-cultural transformations when aligned with sustainable development policies. Furthermore, this study has opened up a fresh perspective in this area by exploring the feasibility of AI applications based on the unique traits and principles of Saudi society and culture. Given the similarities among Islamic nations, particularly in the Middle East, this new perspective could hold significant value in establishing a solid groundwork for AI utilization in nations such as Saudi Arabia. With the rapid advancement of smart city initiatives in Saudi Arabia, propelled by numerous ongoing projects, the comprehensive insight offered by this research into the potential of AI applications could prove crucial for future progress.

Methodology

This narrative review employs a structured four-step approach to analyze literature relevant to its main goals. Given that the topics in this area are still developing and lack integrated literature, a narrative review was deemed appropriate. Such reviews are significant in academic research as they offer a comprehensive overview of existing literature, allowing researchers to contextualize their findings and identify knowledge gaps. Unlike systematic reviews, which adhere to strict methodologies for literature selection, narrative reviews prioritize understanding the topic through extensive reading, fostering deeper insights that facilitate better literature synthesis. They also enhance accessibility for a broader audience, encourage engagement with complex subjects, and stimulate scholarly discourse by challenging existing assumptions and highlighting emerging trends (Orben, 2020).

This study incorporates specific criteria and processes in each step of the narrative review to ensure methodological rigor and transparency.

Step 1: Identification of Socio-Cultural Challenges

The first step involves identifying and extracting the most significant socio-cultural challenges facing Saudi Arabia. This is achieved through conducting a thorough review of existing literature on socio-cultural issues in Saudi Arabia, including recent studies and official documents like Saudi Vision 2030. We used this search string on the All Fields section of Scopus database on 20 July 2024:

> ("socio-cultural challenges" OR "cultural issues" "cultural challenges" OR "social dynamics" OR "social challenges" OR " social issues" OR "social problems" OR "cultural problems") AND ("Saudi Arabia" OR "Kingdom of Saudi Arabia" OR "Saudi")

We found 141 articles and after analyzing and reading the title and abstract of them we found 5 main socio-cultural challenges in Saudi Arabia: Gender equality, Cultural preservation and heritage protection, Increasing Youth Population, Rapid Urbanization, and Climate change.

Step 2: Analysis of SDGs

In the second phase, we designed a search string for each socio-cultural challenge we identified in the previous step. By applying these search strings in the Title section of the Scopus database we identified suitable documents for connecting SDGs to our socio-cultural challenges. The search strings are:

> Gender equality: ("Sustainable Development Goals" OR "SDGs" OR "UN SDGs" OR "2030 Agenda" OR "global goals") AND ("gender equality" OR "gender equity" OR "women's empowerment" OR "gender parity" OR "gender rights" OR "women's rights" OR "gender issues" OR "gender discrimination" OR "female empowerment" OR "equal opportunities for women" OR "gender balance")

Cultural preservation and heritage protection: ("Sustainable Development Goals" OR "SDGs" OR "UN SDGs" OR "2030 Agenda" OR "global goals") AND ("cultural preservation" OR "heritage protection" OR "cultural heritage conservation" OR "cultural sustainability" OR "heritage conservation" OR "cultural safeguarding" OR "cultural legacy protection" OR "preservation of cultural heritage" OR "intangible cultural heritage" OR "tangible cultural heritage")

Increasing Youth Population: : ("Sustainable Development Goals" OR "SDGs" OR "UN SDGs" OR "2030 Agenda" OR "global goals") AND ("increasing youth population" OR "growing youth demographic" OR "youth bulge" OR "youth growth" OR "young population" OR "youth expansion" OR "demographic transition" OR "youth development" OR "population growth among youth" OR "young people")

Rapid Urbanization: ("Sustainable Development Goals" OR "SDGs" OR "UN SDGs" OR "2030 Agenda" OR "global goals") AND ("rapid urbanization" OR "accelerated urban growth" OR "urban expansion" OR "urban sprawl" OR "urban development" OR "fast urbanization" OR "urban migration" OR "urbanization trends" OR "urban population growth")

Climate change: ("Sustainable Development Goals" OR "SDGs" OR "UN SDGs" OR "2030 Agenda" OR "global goals") AND ("climate change" OR "global warming" OR "climatic change" OR "climate crisis" OR "climate variability" OR "climate disruption" OR "environmental change" OR "atmospheric changes" OR "climate impacts")

Step 3: Identification of AI Applications

The third step focuses on identifying and categorizing the most significant AI applications that can be utilized to implement and strengthen the relevant SDGs. In this regard, the global literature is reviewed to find successful AI applications that tackle similar socio-cultural challenges and promote sustainable development. We applied a relevant search string on Title section of Scopus database:

Step 4: Discussion of Results and Recommendations

In the final phase, the study will discuss the findings from the third step, focusing on the effectiveness and necessary conditions for promoting the identified AI applications in the context of Saudi Arabia. This will include an analysis of cultural context considerations, examining how Saudi Arabia's socio-cultural characteristics — such as traditional values and ongoing projects like NEOM — impact the adoption of these AI technologies. Additionally, a needs assessment will be conducted to identify the resources, training, and infrastructure required for effective implementation of the AI applications. Finally, the study will propose actionable policy recommendations aimed at facilitating the integration of AI in promoting relevant SDGs while addressing socio-cultural challenges in Saudi Arabia.

Additionally, this step involves critically evaluating potential barriers to AI implementation, such as data privacy concerns, ethical considerations, and resistance to tech-

nological change. Strategies to mitigate these barriers are also discussed to ensure AI solutions' successful adoption and sustainability.

Inclusion and Exclusion Criteria

To enhance the transparency and reproducibility of the narrative review, specific inclusion and exclusion criteria have been established. The inclusion criteria specify that publications must be from the last five years to ensure the incorporation of recent developments and trends. Only peer-reviewed journal articles, conference papers, and official government reports will be considered. Additionally, studies must focus on AI applications, SDGs, and socio-cultural challenges in Saudi Arabia or comparable contexts in the Global South. Research providing empirical evidence, case studies, or comprehensive reviews relevant to the research goals will also be included. Conversely, the exclusion criteria indicate that publications older than five years will be excluded unless they are seminal works essential for foundational understanding. Studies that do not directly relate to Saudi Arabia's socio-cultural challenges or AI applications will also be omitted. Furthermore, opinion pieces, editorials, and non-peer-reviewed sources lacking substantial evidence or analysis will not be included in the review.

Quality Assessment

To ensure the reliability and validity of the included studies, a quality assessment was conducted based on several criteria. First, relevance was considered, meaning that the study must directly address one or more of the research goals. Methodological rigor was also essential; the study should employ robust and appropriate data collection and analysis methodologies. Additionally, credibility was evaluated by checking whether the study was published in reputable journals or by recognized institutions.

The strength of evidence was another critical factor, requiring that the study provide solid empirical evidence or a comprehensive theoretical analysis. Finally, bias minimization was assessed by examining whether the study demonstrated efforts to minimize bias through transparent reporting and acknowledgment of limitations. Each study included in the review was evaluated against these criteria, and only those meeting the majority of the standards were incorporated into the final analysis.

Addressing Potential Biases

Recognizing the inherent limitations of narrative reviews, this study proactively addresses potential biases through several strategies. First, it diversifies sources by incorporating a variety of materials from different disciplines and perspectives to provide a balanced view. The selection process is also made transparent by clearly documenting the inclusion and exclusion criteria, ensuring a systematic and unbiased approach to selecting studies.

Additionally, the study engages in critical appraisal, actively critiquing the strengths and weaknesses of each included study to avoid over-reliance on any single source or

perspective. Finally, reflexivity is emphasized by acknowledging the researcher's biases and perspectives while striving to maintain objectivity throughout the review process.

Literature review

Saudi Arabia Socio-Cultural Challenges

As the birthplace of Islam, Saudi Arabia is a country steeped in religious and cultural values, and its unique economic, religious, political, and socio-cultural characteristics have made this country an influential country in the Middle East and the world. Nevertheless, Saudi Arabia is modernizing at a high speed, and at the same time, it is facing important socio-cultural challenges. Identifying socio-cultural challenges in a country is very important to promote social justice, preserve cultural heritage, strengthen social cohesion, improve quality of life, and support economic development. By addressing these challenges, countries can strive to create a more just, inclusive, and prosperous society for all (S.K. Yadav, 2023). Therefore, we will discuss the most important recent socio-cultural challenges of Saudi Arabia.

Gender Equality

Researchers consider gender equality in Saudi Arabia as one of the most important socio-cultural challenges in Saudi Arabia. The 2020 Global Gender Gap Index report shows Saudi Arabia's low ranking in terms of gender equality, with women making up only 16% of the workforce. (Alnufaie, Beghum, 2021). Despite some progress in empowering women and changing the dynamics of the work environment, Saudi women still face challenges such as gender segregation and lack of representation in the workforce (Aldossari, Calvard, 2022). The country's cultural climate, under the influence of male guardianship laws and conservative norms, creates an unfriendly environment for women and affects their movement and interaction. (Alasmari, 2023; Alsharif, Ulrich-Schad, 2019). While recent efforts have been made for gender equality and women's empowerment in Saudi Arabia, there is still a long way to go to remove systemic barriers and promote equal opportunities for women in various areas of society (Al-Nasrallah, 2023).

Recent studies emphasize the intersectionality of gender with other socioeconomic factors, highlighting how AI can be leveraged to promote gender equality by enhancing women's access to education, employment, and healthcare (Mishra et al., 2023; Olubiyi et al., 2022). Additionally, AI-driven initiatives can help monitor and enforce gender-related policies, thereby facilitating a more equitable workforce (Baena-Morales et al., 2020).

Cultural Preservation and Heritage Protection

Saudi Arabia is rich in cultural and historical sites, with a diverse population and important cultural and behavioral values. Despite the development of the non-religious tourism industry and the rapid modernization of cities, cultural and heritage protection

remains a challenge. Research highlights the importance of local assessment criteria for cultural heritage values and conservation, focusing on identifying and emphasizing the importance of places through their values (Albaqawy et al., 2023). Initiatives in Saudi Arabia emphasize the preservation of historical heritage, ancient materials, and adaptive reuse methods to maintain social values and promote sustainable conservation approaches (MAZZETTO, 2022). The Saudi National Vision 2030 emphasizes the importance of preserving heritage sites and the local environment to strengthen national identity and Arab values (MAZZETTO, 2022). Remote sensing and geographic information systems are used to produce risk maps for World Heritage sites, helping decision-makers develop proactive conservation programs (Ramadan et al., 2022). Interventions such as university expansion and airport renovations align with the country's goal of achieving a sustainable economy and tourism. More comprehensive policies and strategies are needed to effectively safeguard both tangible and intangible aspects of Saudi Arabia's cultural legacy (Aldegheishem, 2024; Alshehaby, 2024; Altassan, 2023).

AI technologies, notably Computer Vision and Geospatial AI, play a pivotal role in cultural preservation by enabling accurate 3D modeling of heritage sites, monitoring environmental changes, and facilitating virtual tourism (Fomin et al., 2022; Roa, Triana, 2021). These applications not only aid in preserving physical structures but also in maintaining the intangible cultural heritage by documenting and analyzing traditional practices and community interactions (Brown Dr et al., 2019).

Increasing Youth Population

Saudi Arabia's rapidly growing youth population, over 60% under 30, has pressured public services and infrastructure, particularly in urban areas (Alasmari, 2023; Alsharif, Ulrich-Schad, 2019). Prioritizing urban development should focus on urban facilities, urban economy, health, housing, and third places to meet the needs of the young generation. However, this population also presents challenges such as high unemployment rates, potential social unrest, and limited educational and employment opportunities. Issues such as job discrimination, lack of appropriate workforce preparation curriculum, and women's reluctance to join the workforce are prevalent (Al-Otaibi, Mansour, 2021). Vision 2030 aims to address youth unemployment through strategic development initiatives, but progress remains a challenge (Asem et al., 2024). Adolescents in Saudi Arabia face health risks like smoking, traffic accidents, and drug abuse, which can impact their well-being and prospects. A limited understanding of youth education and employment pathways after high school complicates the situation, highlighting the need for comprehensive reforms and support systems. The government must address these challenges to support and benefit from the growing youth population in Saudi Arabia (Algarni et al., 2023; Alqahtany, Aravindakshan, 2022; Khateb, Alkhaibari, 2023).

AI applications, such as Predictive Analytics and Natural Language Processing, can significantly enhance educational outcomes by personalizing learning experiences and predicting student performance, thereby addressing educational disparities (Milicevic et

al., 2024; Okulich-Kazarin et al., 2023). Furthermore, AI-driven job matching platforms and career counseling tools can help reduce youth unemployment by aligning educational outcomes with labor market demands (Dawana et al., 2024; Yue et al., 2021).

Rapid Urbanization

Urban development in Saudi Arabia, particularly in the Al-Ahsa Urban Area, has led to significant demographic challenges and environmental consequences (Alqahtany, 2023). The rapid growth has resulted in pollution, water resource degradation, urban heat island effect, and urban sprawl, which negatively impact the environment and public health (Almulhim, Cobbinah, 2023). Unsustainable urbanization practices, particularly in Riyadh and Jeddah, contradict sustainability principles, emphasizing the need for sustainable urban policies. The trade-off between heritage site protection and increased urbanization complicates cultural heritage management (Alqahtany, Aravindakshan, 2022). The rapid growth of urban populations has led to a housing supply gap, hindering the government's goal of increasing home ownership to 70% by 2030 (Alhamoudi, 2024). Traffic congestion and accidents in cities like Abha and Bisha are also a concern. The rapid urbanization and ambitious development policies in cities like "Buraydah" have disrupted traditional environments, leading to a loss of local identity. Key challenges include data management infrastructure, construction industry regulations, uncertain housing delivery policies, and resistance to change (Abuhasel, 2023). In this regard, the Saudi government has launched initiatives like the Saudi Vision 2030 to increase home ownership and promote sustainable urban development. However, successful implementation requires a comprehensive strategy that addresses the various urbanization challenges facing the country (Al-Ansi et al., 2023; Alhamoudi, 2024).

AI-driven smart city technologies, including IoT, machine learning, and Geospatial AI, are essential in managing urban growth sustainably (Karpov et al., 2023; Kumar et al., 2023). These technologies enable efficient resource allocation, traffic management, and environmental monitoring, thereby mitigating the adverse effects of rapid urbanization (Maroju et al., 2023; Pandey et al., 2024). Additionally, AI can facilitate the integration of green spaces and optimize energy consumption in urban infrastructure, aligning with sustainability goals (Meng et al., 2018; Tawfik et al., 2024).

Climate Change

Saudi Arabia faces significant challenges due to climate change, particularly in Dammam. Despite 90% of respondents being aware of climate change, 33% were unaware of its causes and impacts. This highlights the need for increased public awareness and education on climate change in the country (Almulhim, 2021). Climate change is also linked to geopolitical risks in Saudi Arabia, with a solid bidirectional causal relationship between the country's geopolitical risk index and critical climate change factors (Dhifaoui et al., 2023). This suggests that climate change is exacerbating regional instability. Climate change also threatens Saudi Arabia's food system security, with factors like temperature

increases, greenhouse gas emissions, population growth, and economic development contributing to vulnerabilities (Al Jaafreh, Allouzi, 2023; Almulhim, 2021; Dhifaoui et al., 2023). To ensure long-term food security, policymakers must address these complex challenges. A multi-faceted approach involving public education, risk management frameworks, and policy interventions is required. Sustainable urban planning, capacity-building programs for farmers, and climate change mitigation and adaptation policies are essential for a resilient national food system and safe roads for all (Abubakar, Dano, 2020; Azeem, Alhafi Alotaibi, 2023) .

AI applications, such as Predictive Analytics, Machine Learning, and Geospatial AI, are instrumental in climate change mitigation and adaptation strategies (Dube et al., 2024; Hamdan et al., 2024). These technologies enable accurate climate modeling, early warning systems, and optimized resource management, which are crucial for addressing the multifaceted impacts of climate change (Sen et al., 2021; Sen et al., 2021). Additionally, AI can enhance agricultural practices by predicting weather patterns, optimizing irrigation systems, improving crop resilience, and supporting food security initiatives (Elufioye et al., 2024; Sakapaji, Puthenkalam, 2023).

Related SDGs and AIs

The SDGs integrate economic, social, and environmental concerns, emphasizing human rights and good governance (Kannengießer, 2023). They are essential in promoting global health and aligning with social entrepreneurship models to address vulnerable populations (Yi, 2023). Implementing the SDGs positively impacts society, including economic development and competitiveness, and can help reduce socio-cultural challenges in Saudi Arabia (Del-Aguila-Arcentales et al., 2022; Surmeli et al., 2022). As Table 1 shows, by reviewing the relevant literature, SDGs that were directly related to solving these challenges were identified for each of the five main challenges. It was found that 10 of the 17 SDGs are directly related to the socio-cultural challenges of Saudi Arabia. On the other hand, a deeper examination of the research showed that for the implementation of the SDGs, AI could be a suitable tool for providing innovative solutions and data-based insights due to its efficiency in the implementation and promotion of the SDGs (Ziesche et al., 2023). Despite the fragmented nature of AI implementation, techniques such as machine learning and explainable AI enable policymakers to measure SDG indicators and make informed decisions effectively (Tuğaç, 2023). As Table 1 shows, as a result of further literature studies and valuable experiences in AI and SDGs, there are direct connections between the two fields. Table 1 identifies key AI applications and models to implement and strengthen the ten main SDGs, which are crucial for Saudi policymakers and designers considering the rapid development of smart cities. The AI models and methods are categorized into six broad areas:

- Predictive Analytics and Forecasting: Includes forecasting analytics and classification tasks for predicting future outcomes or classifying data (Nagahisarchoghaei et al., 2023; Zhang et al., 2020).

- Optimization and Decision Support: Covers optimization tasks, recommendation systems, and reinforcement learning applications for decision-making and optimization (Ma et al., 2023).
- Natural Language Processing: Focuses on understanding, interpreting, and generating human language (İpek et al., 2023).
- Computer Vision: Involves processing and analyzing visual data (Karpov et al., 2023).
- Generative AI: Includes applications that create new content (Mohammadi & Kalhor, 2021).
- Geospatial AI: Focuses on analyzing and interpreting geographic and spatial data (Karpov et al., 2023).

Table 1. Related SDGs, AI, models and methods

SDG Area	AI Applications	Key AI Models and Methods
SDG 4: Quality Education	• Enhance learning outcomes and personalize learning experiences (Milicevic et al., 2024) • Develop future skills using intelligent tutoring systems and advanced methods (Okulich-Kazarin et al., 2023)	• Natural Language Processing (Zhang, 2024) • Generative AI (gshayish, 2023) • Large Language Models (Arruda, 2024) • Computer Vision (Okulich-Kazarin, 2023) • Recommendation Systems (Okulich-Kazarin, 2023)
SDG 5: Gender Equality	• Promote women's political participation and address gender violence (Mishra et al., 2023) • Develop machine learning models for gender equality policies and governance structures (Olubiyi et al., 2022) • Empower women through ICT and wearable technologies for breastfeeding support (Shah & Krishnan, 2024) • Address gender stereotypes in educational content and media (Baena-Morales et al., 2020)	• Natural Language Processing (Azmi, 2020) • Sentiment Analysis (Baena-Morales et al., 2020) • Forecasting Analytics (Olubiyi et al., 2022) • Recommendation Systems (Nurmila et al., 2021) • Reinforcement Learning (Kathambi & Obiero, 2022) • Computer Vision (Mishra & Mishra, 2023) • Large Language Models (Shah & Krishnan, 2023) • Generative AI (Baena-Morales et al., 2020)
SDG 6: Clean water and sanitation	• Implement smart water management systems and enhance water governance (Maroju et al., 2023) • Improve wastewater treatment and water desalination processes (S. Pandey et al., 2022) • Monitor and assess water quality and sanitation facilities (Maroju et al., 2023)	• Optimization models (Maroju, 2023;) • Classification models (Maroju, 2023) • Large language models (Maroju, 2023) • Forecasting Analytics (Maroju, 2023) • Computer Vision (Maroju, 2023) • Generative AI (Pandey et al., 2022 • Reinforcement Learning (Saboori & Mehrjerdi, 2022)

SDG 7: Affordable and clean energy	• Optimize energy consumption in smart homes and manage energy in smart green ports (Tawfik et al., 2024) • Enhance energy forecasting and optimization in buildings and grids (Meng et al., 2018) • Improve power management in IoT devices (A. K. Pandey et al., 2024)	• Optimization models (Tawfik, 2024(• Forecasting Analytics (Kumar et al., 2023)
SDG 8: Decent work and economic growth	• Analyze labor markets and economic growth (Yu et al., 2019) • Enhance productivity in manufacturing and secure IoT environments (Yue et al., 2021) • Utilize Augmented Reality and AI for interactive learning experiences (Dawana et al., 2024)	• Forecasting Analytics (Yue et al., 2021) • Classification models (Yue et al., 2021)
SDG 11: Sustainable cities and communities	• Develop transport operations, urban planning, and smart city initiatives (Marji et al., 2024) • Conduct sentiment analysis towards urban initiatives and monitor infrastructure (Ametepey et al., 2024) • Promote green AI practices and community development (Arruda & Arruda, 2024) • Address ethical considerations in AI use (Theodorou et al., 2022)	• Forecasting Analytics (Marji et al., 2024) • Classification models (Ametepey et al., 2024) • optimization models (Cirianni et al., 2023) • natural language processing (Ametepey et al., 2024) • Sentiment analysis (Ametepey et al., 2024) • Computer vision (Ametepey et al., 2024) • Generative AI (Hasas et al., 2024)
SDG 12: Responsible consumption and production	• Forecast consumer trends and optimize production processes (Hannan et al., 2021) • Categorize products by environmental impact and understand consumer sentiments (Matsui et al., 2022) • Enhance supply chain management and design sustainable packaging solutions (Kurnia et al., 2023)	• Forecasting Analytics (Hannan et al., 2021) • Classification models (Bjola, 2022) • Natural language processing (Matsui et al., 2022) • sentiment analysis (Matsui et al., 2022) • Optimization models (Kurnia et al., 2023) • Computer vision (Okulich-Kazarin et al., 2023) • Generative AI (Pigola et al., 2021)
SDG 13: Climate action	• Mitigate climate change and develop sustainable solutions (Hamdan et al., 2024) • Forecast solar photovoltaic power and improve renewable energy efficiency (Iheanetu, 2022) • Enhance climate resilience, weather forecasting, and water resources management (Dube et al., 2024) • Promote climate-resilient agricultural practices and food security (Sakapaji & Puthenkalam, 2023)	• Forecasting Analytics (Hamdan et al., 2024) • classification models (Sen et al., 2021) • optimization models (Sakapaji & Puthenkalam, 2023)

SDG 14: Life below water	• Predict marine conservation trends and identify risks (Joel et al., 2024) • Conserve marine biodiversity and forecast marine traffic dynamics (Molina-Molina et al., 2021) • Predict ocean dynamics and extreme events (Dong et al., 2022)	• Forecasting Analytics (Nzeako et al., 2024) • Geo AI (Molina-Molina et al., 2021)
SDG 15: Life on land	• Monitor the environment, conserve biodiversity, and detect land cover changes (Chisom et al., 2024) • Optimize agricultural supply chains and forecast species abundances (Elufioye et al., 2024) • Predict habitat quality under climate change scenarios and manage inventories (Nzeako et al., 2024)	• Geo AI (Chisom et al., 2024) • Forecasting Analytics(Nzeako et al., 2024) • classification models (Limberger et al., 2024) • natural language processing (Kass et al., 2022) • computer vision (Kass et al., 2022)

AI's role in supporting SDGs extends beyond mere implementation; it actively contributes to formulating strategies that align with sustainable and inclusive growth. For instance, Natural Language Processing (NLP) can be utilized to analyze public sentiment and feedback on gender policies, thereby facilitating more effective and responsive governance (Baena-Morales et al., 2020). Similarly, Computer Vision technologies can monitor and preserve cultural heritage sites by detecting real-time structural changes or potential threats, ensuring timely interventions (Fomin et al., 2022).Furthermore, the integration of AI in urban planning through Geospatial AI allows for the creation of more resilient and adaptable infrastructure, which is essential for managing the challenges posed by rapid urbanization and climate change (Karpov et al., 2023; Maroju et al., 2023). These AI-driven solutions address immediate socio-cultural challenges and contribute to long-term sustainable development by fostering innovation, efficiency, and inclusivity in Saudi Arabia's smart cities.

Additionally, it is important to consider the ethical implications and sustainability of AI applications. AI's water footprint, which refers to the amount of water consumed during the training and operation of AI models, poses significant environmental concerns, especially in water-scarce regions like Saudi Arabia (Chen et al., 2023). Addressing this issue requires the development of more efficient algorithms and adopting green AI practices to minimize environmental impact while maximizing the benefits of AI technologies (Mashhood et al., 2023).

Moreover, recent studies have highlighted the importance of incorporating local cultural contexts into AI applications to ensure their relevance and effectiveness. By tailoring AI solutions to Saudi Arabia's unique socio-cultural dynamics, policymakers can enhance the acceptance and success of these technologies in achieving the SDGs (Olubiyi et al., 2022; Samarin, Al-Asfour, 2023).

In summary, the interplay between AI and SDGs in the context of Saudi Arabia's socio-cultural challenges is multifaceted and dynamic. The following sections will delve deeper into specific AI applications and their contributions to each identified SDG, com-

prehensively analyzing how AI can drive sustainable and inclusive development in Saudi Arabia's smart cities.

Results and Discussion

In general, Saudi Arabia is struggling with various socio-cultural challenges, and the sustainability of the society depends on solving these challenges. This is also emphasized in the Saudi 2030 Vision document. Therefore, this research *provides an overview* of the most important challenges in Saudi Arabia. This study first identifies the most likely SDGs to solve these problems and then the most important AI applications *used to implement and promote these SDGs*.

Saudi Arabia is implementing ambitious smart city projects such as NEOM, which can have trans-regional impacts. On the other hand, AI is considered an important tool in smart cities. Integrating AI technologies into various urban systems is a key factor in developing smarter, more efficient, and livable smart cities (Napolitano, 2023; Rebahi et al., 2023). As a result, this integration should align with the socio-cultural challenges and characteristics of Saudi Arabia to drive sustainability and enhance the quality of life. Nevertheless, we see that the Neom project, which is the most prominent project of the smart city, has followed the emphasis on solving some socio-cultural challenges to a lesser extent. In general, this project has emphasized goals such as biophilic design (Alhefnawi, 2022), technology (AlDoaies, Almagwashi, 2018), geothermal and renewable energy (Aboud et al., 2023), urban investment bonds, and geological and environmental considerations (Chang, 2024), which can be in line with the challenges of rapid urbanization, climate change, and increasing youth population. However, the two difficulties of gender equality and Cultural preservation and heritage protection do not play a central role in the Neom project. At the same time, there are some criticisms of this project. For example, the researchers believe the NEOM site is located near an active geological fault, which is risky. The researchers also believe disruption to local ecosystems and habitats is possible due to the large-scale construction and development of NEOM (Alhefnawi, 2022; Bashir, Alsalman, 2023). Therefore, despite the concrete emphasis of Saudi 2030 vision on socio-cultural challenges, the priority of some of these challenges cannot be identified, at least in the NEOM project. There is a need to review the tools, policies, and technologies in NEOM and other Saudi smart city projects, which this research saw in the appropriate use of AI. In this regard, Figure 1 was designed, which shows the close relationship between Saudi socio-cultural challenges with SDGs and AI applications, and this can be considered an important perspective for researchers and policymakers in the field of smart cities in Saudi Arabia.

Gender Equality

Gender equality is a significant socio-cultural challenge in Saudi Arabia, aligning with SDG 5. AI applications have been developed to improve women's economic and political

participation (Mishra et al., 2023), gender equality policies (Olubiyi et al., 2022), and address gender stereotypes in content and educational media (Baena-Morales et al., 2020). These applications use Natural Language, Processing, Predictive Analytics, Forecasting, Optimization and Decision Support, Computer Vision, and Generative AI models. In smart cities, AI should be designed and implemented with the help of women to maximize efficiency in gender equality. AI-driven initiatives can facilitate monitoring gender-related policies and analyzing workforce data to identify and address disparities (Baena-Morales et al., 2020). Additionally, Natural Language Processing can be employed to analyze public discourse and media content to identify and mitigate gender biases, thereby fostering a more inclusive environment (Mishra et al., 2023).

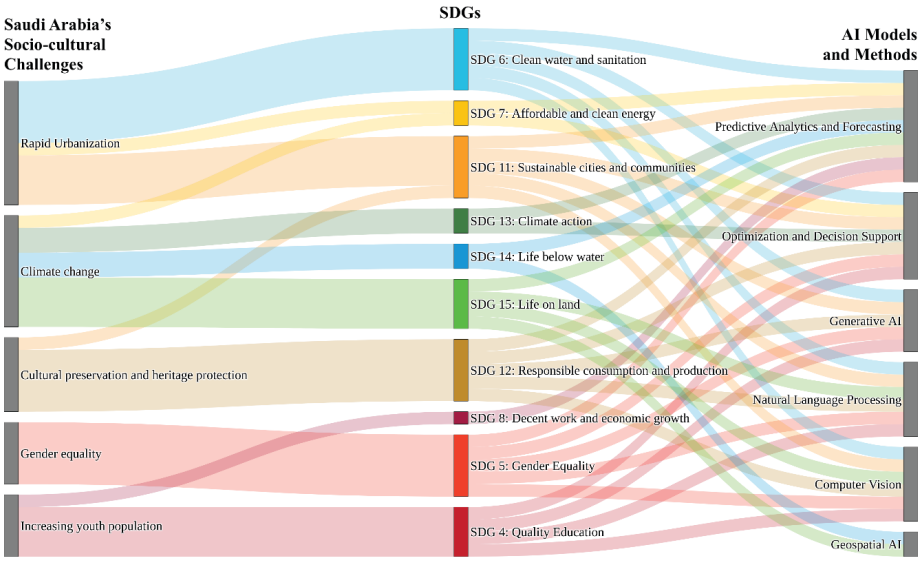


Figure 1. Final overview of SDGs in relation to Saudi Arabia's socio-cultural challenges with AI models and methods

Cultural Preservation and Heritage Protection

Related to SDG 11, the preservation and protection of cultural heritage are also crucial, encompassing not only ancient architecture and urban planning but also the identity, cultural heritage, and social values of Saudi Arabia. AI models can be used to create accurate 3D models of heritage sites and artifacts, monitor and document them (Roa, Triana, 2021), track changes in their condition (Brown Dr et al., 2019), and analyze environmental data to predict potential threats (Yildirim et al., 2019). Virtual reality and augmented reality can improve tourism by optimizing visitor flow, monitoring crowd density, and providing personalized recommendations (Barbosa et al., 2022). In SDG 12, AI applica-

tions can be used for heritage conservation, including digital documentation, condition monitoring, restoration, and interaction with visitors (Fomin et al., 2022). These capabilities contribute to responsible consumption and production of natural resources, reduce waste, and promote sustainable practices. However, there are negative points to consider when preserving and protecting cultural heritage using AI. It is essential to ensure that AI systems respect cultural sensitivities, intellectual property rights, and autonomy of local communities and that AI-based cultural preservation initiatives are accessible and inclusive for diverse communities. As Saudi Arabia is a significant pilgrimage destination for Muslims and a form of Islamic cultural crystallization, it is crucial to preserve the authenticity and accurate representation of cultural practices and identities, avoiding simplification or misinterpretation in AI applications and smart cities.

AI technologies such as Computer Vision and Geospatial AI play a pivotal role in cultural preservation by enabling the creation of detailed 3D models of heritage sites, facilitating real-time monitoring, and predicting potential risks from environmental changes (Fomin et al., 2022; Roa, Triana, 2021). These technologies ensure that cultural heritage is accurately documented and preserved for future generations while enhancing tourism through immersive virtual experiences (Barbosa et al., 2022).

Increasing Youth Population

The increasing youth population in Saudi Arabia requires a focus on promoting SDGs 4 and 8 in smart cities. AI applications can accelerate SDG 4 by improving learning outcomes and personalizing experiences, enabling the development of future skills (Milicevic et al., 2024; Okulich-Kazarin et al., 2023). The young society also needs a dynamic economy and job opportunities aligned with SDG 8. Predictive Analytics and Forecasting models can analyze labor market and economic growth, increase productivity in production, and create safe environments for the Internet of Things (Dawana et al., 2024; Yu et al., 2019; Yue et al., 2021). AI-powered educational platforms can personalize learning experiences to better cater to the diverse needs of the youth, thereby enhancing educational outcomes and reducing dropout rates (Milicevic et al., 2024). Furthermore, AI-driven job matching systems can bridge the gap between graduates and employment opportunities, addressing high unemployment rates and aligning workforce skills with market demands (Yue et al., 2021).

Rapid urbanization

Rapid urbanization is a comprehensive challenge in Saudi Arabia, requiring sustainable city planning, water and urban waste management, and access to clean energy. AI applications can play a crucial role in promoting SDG 6 by implementing smart water management systems, strengthening water governance, improving wastewater treatment and water desalination processes, and monitoring and evaluating water and sanitary facilities (Maroju et al., 2023; Mraz et al., 2021; S. Pandey et al., 2022). AI applications can also op-

timize energy consumption in smart homes, smart green ports, buildings and networks, and Internet of Things devices, promoting and implementing SDG 7 (Kumar et al., 2023; Meng et al., 2018; A. K. Pandey et al., 2024; Tawfik et al., 2024). SDG 11 is the most important SDG related to the challenge of rapid urbanization in Saudi Arabia and should be taken seriously. Successful AI experiences at the global level can be considered a suitable tool for creating sustainable smart cities in Saudi Arabia.

AI-driven smart water management systems utilize machine learning algorithms to predict water demand, detect leaks, and optimize distribution networks, ensuring efficient water resource use (Maroju et al., 2023). Similarly, AI applications in energy management can optimize consumption patterns in smart homes and buildings, significantly reducing energy wastage and promoting the use of renewable energy sources (Kumar et al., 2023). These AI solutions are essential for managing the complexities of rapid urbanization and ensuring sustainable growth in Saudi Arabia's smart cities.

Climate Change

The climate change challenge in Saudi Arabia is closely linked to SDGs 7, 13, 14, and 15. The NEOM project emphasizes the importance of overcoming this challenge for the future of smart cities in Saudi Arabia. AI applications have proven successful in addressing SDG 13 by reducing climate change, creating sustainable solutions, forecasting solar photovoltaic power, improving renewable energy efficiency, increasing climate flexibility, weather forecasting, water resource management, and promoting agricultural practices resistant to climate and food security (Dube et al., 2024; Hamdan et al., 2024; Iheanetu, 2022; Sakapaji, Puthenkalam, 2023; Sen et al., 2021). Saudi Arabia's wide border with the Arabian Gulf and the Red Sea is also related to SDG 14, and its strengthening is crucial. AI applications can help promote SDG 14 by predicting sea protection trends, preserving marine biodiversity, predicting marine traffic dynamics, and predicting ocean dynamics and extreme events (Dong et al., 2022; Joel et al., 2024; Molina-Molina et al., 2021; Mooney et al., 2020). AI applications can also help improve and implement SDG 15 by monitoring the environment, preserving biodiversity, identifying land cover changes, optimizing the agricultural supply chain, predicting species abundance, and managing inventory (Chisom et al., 2024; Elufioye et al., 2024; Limberger et al., 2024; Nzeako et al., 2024; Requena-Mullor et al., 2017). Given Saudi Arabia's special and sensitive climatic conditions, future smart cities can use AI to effectively fight climate change challenges and achieve SDGs 7, 13, 14, and 15.

AI technologies such as Predictive Analytics and Machine Learning enable accurate climate modeling and weather forecasting, which is critical for developing effective climate change mitigation and adaptation strategies (Sen et al., 2021; Hamdan et al., 2024). Geospatial AI assists in monitoring environmental changes and predicting extreme weather events, thereby enhancing the resilience of infrastructure and communities (Dong et al., 2022; Molina-Molina et al., 2021). Additionally, AI-driven agricultural technologies can optimize irrigation, improve crop yields, and promote sustainable farming practices, ad-

dressing both food security and environmental sustainability (Sakapaji, Puthenkalam, 2023; Elufioye et al., 2024). Although limited, other studies have also attempted identify suitable AI technologies for addressing the socio-cultural challenges of Saudi Arabia.

Conclusion

The study explores the utilization of AI to address socio-cultural challenges in Saudi Arabia by promoting SDGs. A narrative review method was used in four steps to achieve this goal. Saudi Arabia's most important socio-cultural challenges were gender equality, increasing youth population, rapid urbanization, and climate change. Ten SDGs were identified as critical for addressing Saudi Arabia's most important socio-cultural challenges using proper AI applications, models, and methods.

Therefore, this study demonstrates that AI applications can play a crucial role in addressing sociocultural challenges by implementing SDGs in various ways based on extensive literature reviews and successful global case studies. However, this research argues that for the correct use of AI in Saudi Arabia, it is necessary to prioritize social, religious, and cultural characteristics and values and that AI should be highly compatible with Saudi society. AI can enhance sustainability in Saudi Arabia's smart cities and contribute to achieving the SDGs, so investing in developing and implementing AI-based solutions is essential to address the identified challenges and foster sustainable development. Policymakers, urban planners, and technology experts must collaborate to leverage AI's potential for maximizing positive societal impact. By doing so, Saudi Arabia can create smart, sustainable communities that improve its residents' and future generations' quality of life.

To translate these findings into actionable strategies, several practical recommendations are proposed. First, establishing national frameworks is essential to guide the integration of AI technologies in smart city projects, ensuring alignment with SDGs and socio-cultural values. Additionally, fostering partnerships among government agencies, private sector entities, academic institutions, and local communities will drive inclusive and culturally sensitive AI initiatives. Implementing educational programs and training workshops is also crucial for building AI literacy and expertise among the youth and workforce, thereby supporting SDG 4 (Quality Education) and SDG 8 (Decent Work and Economic Growth). Furthermore, developing and enforcing ethical guidelines for AI deployment will address data privacy, security, and ethical concerns, ensuring that AI applications respect cultural and societal norms. Lastly, providing funding for research and development in AI technologies tailored to Saudi Arabia's unique socio-cultural context will foster innovation and sustainable solutions.

Despite the significant potential of AI, this study acknowledges several limitations. One notable challenge is the implementation of AI solutions, which may encounter barriers such as inadequate infrastructure, limited access to quality data, and resistance to technological change. Additionally, data privacy and ethical concerns are paramount; ensuring the privacy and security of data used in AI applications is particularly critical in culturally sensitive contexts. Furthermore, the generalizability of the findings is a limita-

tion to consider. While the study focuses on Saudi Arabia, the insights gained may not be applicable to other regions in the Global South due to differing socio-cultural dynamics.

Future research should focus on several key areas to address existing limitations and further explore the role of AI in sustainable development. One area of investigation is the implementation barriers that hinder AI adoption in Saudi Arabia's smart cities. Conducting empirical studies to identify and analyze these obstacles will be crucial for developing effective strategies to overcome them. Another important area is the exploration of ethical frameworks. There is a need to develop comprehensive ethical guidelines that address data privacy, security, and the responsible use of AI in culturally diverse settings. This will help ensure that AI technologies are implemented in ways that respect individual rights and societal norms. Assessing the long-term impacts of AI integration is also vital. Evaluating the social, economic, and environmental consequences of AI in smart cities will help ensure that its benefits are sustained and aligned with SDGs. Also, promoting inclusive AI development is essential. Research should focus on methods to involve diverse stakeholders, including marginalized communities, in the design and implementation of AI technologies to ensure inclusivity and equity. This collaborative approach will enhance the relevance and effectiveness of AI initiatives within various socio-cultural contexts.

In conclusion, this study underscores AI's transformative potential in advancing sustainable development and addressing socio-cultural challenges in Saudi Arabia's smart cities. By prioritizing culturally aligned AI applications and fostering collaborative efforts among key stakeholders, Saudi Arabia can harness AI to achieve its Vision 2030 goals, creating resilient and inclusive communities for future generations.

References

Aboud E., Alqahtan, F., Abdulfarraj M., Abraham E., El-Masry N., Osman H. (2023) Geothermal imaging of the Saudi cross-border city of NEOM deduced from magnetic data. *Sustainability*, vol. 15, no 5, pp. 4549.

Abubakar I. R., Aina Y. A. (2018) Achieving sustainable cities in Saudi Arabia: Juggling the competing urbanization challenges. *E-planning and collaboration: Concepts, methodologies, tools, and applications,* IGI Global, pp. 234–255.

Abubakar I. R., Dano U. L. (2020) Sustainable urban planning strategies for mitigating climate change in Saudi Arabia. *Environment, Development and Sustainability*, vol. 22, no 6, pp. 5129–5152.

Abuhasel K. A. (2023) A GIS Approach for Analysis of Traffic Accident Hotspots in Abha and Bisha Cities, Saudi Arabia. *Sustainability*, vol. *15, no* 19, pp. 14112.

Adibhesami M. A., Karimi H., Sepehri B. (2024) *Optimizing Urban Design for Pandemics Using Reinforcement Learning and Multi-objective Optimization*. https://doi.org/10.1007/978-981-99-9014-6_5

Al Jaafreh M. B., Allouzi M. A. (2023) Smart Cities Adoption in Saudi Arabia: A Comprehensive Review and Future Drivers. *British Journal of Multidisciplinary and Advanced Studies*, vol. 4, no 5, pp. 20–39.

Al-Ansi N. A., Uddin B., Alhrabi A., Wahid J. (2023) Impacts of Urban Growth Policy on Loss of Identity in Expanding Saudi Cities: A Case Study of Buraydah. *International Journal of Sustainable Development & Planning*, vol. 18, no 10.

Alasmari A. A. (2023) Challenges and social adaptation of international students in Saudi Arabia. *Heliyon*, vol. 9, no 5.

Albaqawy G. A., Alnaim M. M., Bay M. A., Touahmia M. (2023) Assessment of Saudi Arabia's Classification and Selection Criteria for Heritage Sites: A Case Study of Barzan Heritage Area in Hail City. *Sustainability*, vol. 15, no 2, p. 1015.

Aldegheishem A. (2024) Assessing urban sustainability in Saudi Arabia: an empirical evidence from Al-Medina Al-Munawwarah. *Environmental Research Communications*, vol. 6, no 5, p. 055023.

AlDoaies B. H., Almagwashi H. (2018) Exploitation of the promising technology: using blockchain to enhance the security of IoT. *2018 21st Saudi Computer Society National Computer Conference (NCC)*, pp. 1–6.

Aldossari M., Calvard T. (2022) The politics and ethics of resistance, feminism and gender equality in Saudi Arabian organizations. *Journal of Business Ethics*, no 181(4), pp. 873–890.

Algarni M. A., Algarni A. A. M., Alqarni W. A., Alqassim A. Y. (2023) Knowledge and attitude of the general population in Saudi Arabia toward weight management medications (WMMs): a cross-sectional study. *Cureus*, vol. 15, no 8.

Alhamoudi A. M. (2024). Enabling Industry 4.0 Technologies to Drive Sustainable Housing Delivery across the Housing Supply Value Chain in Saudi Arabia: Challenges and Prospects. *Sustainability*, *16*(13), 5413.

Alhefnawi M. A. M. (2022) Integrating the biophilia physiognomies in the context of Neom smart city in Saudi Arabia. *Acta Scientiarum Polonorum Administratio Locorum*, vol. 21, no 2, pp. 159–171.

Alismail S., Faridi M. R. (2021) Harnessing the power of Irtiqaa: An initiative inspired by vision 2030, Kingdom of Saudi Arabia. *Academy of Strategic Management Journal*, vol. 20, no 1, pp. 1–10.

Almajed O. S., Aljouie A., Alghamdi R., Alabdulwahab F. N., Laheq M. T. (2024) Transforming dental care in Saudi Arabia: challenges and opportunities. *Cureus*, vol. 16, no 2.

Almulhim A. I. (2021) Public knowledge and perception of climate change and global warming in the context of environmental challenges and policies in Saudi Arabia. *WIT Transactions on Ecology and the Environment*, no 253, pp. 577–589.

Almulhim A. I., Cobbinah P. B. (2023) Urbanization-environment conundrum: an invitation to sustainable development in Saudi Arabian cities. *International Journal of Sustainable Development & World Ecology*, vol. 30, no 4, pp. 359–373.

Al-Nasrallah W. (2023) The decade long story of gender equality and female empowerment: a case study of corporate disclosures in Saudi Arabia. *Sustainability Accounting, Management and Policy Journal*, vol. 14, no 1, pp. 216–241.

Alnasser A., Musallat N. (2022). Food sustainability knowledge among Saudis: Towards the goals of Saudi Vision 2030. *Sustainability*, vol. 14, no 18, pp. 11398.

Alnufaie H., Beghum S. (2021) Emerging pattern of women empowerment in Saudi Arabia. *Research Review International Journal of Multidisciplinary*, vol. 6, no 5, pp. 21–25.

Alotaibi B. S., Elnaklah R., Agboola O. P., Abuhussain M. A., Tunay M., Dodo Y. A., Maghrabi A., Alyami M. (2024) Enhancing Najran's sustainable smart city development in the face of urbanization challenges in Saudi-Arabia. *Journal of Asian Architecture and Building Engineering*, pp. 1–31.

Al-Otaibi N. M., Mansour D. S. B. M. (2021) Obstacles to Practicing Responsiveness to Intervention in Primary Schools by Learning Disabilities Teachers in Jubail, Saudi Arabia. *Educational Research and Innovation Journal*, vol. 1, no 1, pp. 116–139.

Alotaibi N. S., Alshehri A. H. (2023) Prospers and obstacles in using artificial intelligence in Saudi Arabia higher education institutions — The potential of AI-based learning outcomes. *Sustainability*, vol. 15, no 13, p. 10723.

Alqahtany A. (2023) GIS-based assessment of land use for predicting increase in settlements in Al Ahsa Metropolitan Area, Saudi Arabia for the year 2032. *Alexandria Engineering Journal*, no 62, pp. 269–277.

Alqahtany A., Aravindakshan S. (2022) Urbanization in Saudi Arabia and sustainability challenges of cities and heritage sites: Heuristical insights. *Journal of Cultural Heritage Management and Sustainable Development*, vol. 12, no 4, pp. 408–425.

Alsharif S., Ulrich-Schad J. D. (2019) Unemployed Women in Rural Saudi Arabia: Gender, Mobility and Social Challenges. *Dirasat: Human and Social Sciences*, vol. 46, no 4.

Alshehaby F. (2024) Assessing the Legal Protection of Intangible Cultural Heritage in Saudi Arabia: A Critical Analysis in the Context of the 2003 UNESCO Convention. *Laws*, vol. 13, no 2, p. 13.

Altassan A. (2023) Sustainability of heritage villages through eco-tourism investment (case study: Al-Khabra Village, Saudi Arabia). *Sustainability*, vol. 15, no 9, p. 7172.

Ametepey S. O., Aigbavboa C., Thwala W. D., Addy H. (2024) The Impact of AI in Sustainable Development Goal Implementation: A Delphi Study. *Sustainability*, vol. 16, no 9, p. 3858.

Arruda E. P., Arruda D. P. (2024) Artificial intelligence for SDG 4 of the 2030 agenda: Transforming education to achieve quality, equality, and inclusion. *Sustainable Economies*, vol. 2, no 2, p. 34.

Asem A., Mohammad A. A., Ziyad I. A. (2024) Navigating Digital Transformation in Alignment with Vision 2030: A Review of Organizational Strategies, Innovations, and Implications in Saudi Arabia. *Journal of Knowledge Learning and Science Technology*, vol. 3, no 2, pp. 21–29.

Azeem M. I., Alhafi Alotaibi B. (2023) Farmers' beliefs and concerns about climate change, and their adaptation behavior to combat climate change in Saudi Arabia. *PLoS One*, vol. 18, no 1, p. e0280838.

Baena-Morales S., Jerez-Mayorga D., Fernández-González F. T., López-Morales J. (2020) The use of a cooperative-learning activity with university students: A gender experience. *Sustainability*, vol. 12, no 21, p. 9292.

Balabel A., Alwetaishi M. (2021) Towards sustainable residential buildings in Saudi Arabia according to the conceptual framework of "Mostadam" rating system and vision 2030. *Sustainability*, vol. 13, no 2, p. 793.

Barbosa M. R. de V, Vieira A. O., Peixoto A., Canhos D. A. L., Stehmann J., Menezes M., Maia L. (2022) Building Networks to Promote Knowledge of Brazil's Biodiversity: The experience of the INCT-Virtual Herbarium. *Biodiversity Information Science and Standards*, no 6, p. e91462.

Bashir B., Alsalman A. (2023) Morpho-Hydrological Analysis and Preliminary Flash Flood Hazard Mapping of Neom City, Northwestern Saudi Arabia, Using Geospatial Techniques. *Sustainability*, vol. 16, no 1, p. 23.

Bjola C. (2022) AI for development: Implications for theory and practice. *Oxford Development Studies*, vol. 50, no 1, pp. 78–90.

Brown Dr S. H., Mitchell Dr N., Yildirim Dr E. A., Buckley AM K., Ortsin G. (2019) *Panel 4 Sustainability and Rural Landscapes: CultureNature-based solutions*.

Chang M. (2024) Urban Investment Bond and Green Financial Innovation: Literature Review and Case Studies. *Advances in Economics and Management Research*, vol. 10, no 1, p. 286.

Chen Z., Wu M., Chan A., Li X., Ong Y.-S. (2023) Survey on AI sustainability: emerging trends on learning algorithms and research challenges. *IEEE Computational Intelligence Magazine*, vol. 18, no 2, pp. 60–77.

Chisom O. N., Biu P. W., Umoh A. A., Obaedo B. O., Adegbite A. O., Abatan A. (2024) Reviewing the role of AI in environmental monitoring and conservation: A data-driven revolution for our planet. *World Journal of Advanced Research and Reviews*, vol. 21, no 1, pp. 161–171.

Cirianni F. M. M., Comi A., Quattrone A. (2023) Mobility control centre and artificial intelligence for sustainable urban districts. *Information*, vol. 14, no 10, p. 581.

DARRAJ E. K. (2019) *Build Operate Transfer contracts and how its role is determined to meet Sustainable Development goals in Saudi Arabia According to Saudi Vision 2030*.

Dawana I. R., Prahani B. K., Hariyono E., Ghofur M. A., Wibowo F. C., Bunyamin M. A. H. (2024) Utilisation of augmented reality technology in physics education: A bibliometric analysis and its impact on Sustainable Development Goals (SDGs). *E3S Web of Conferences*, no 513, p. 04006.

Del-Aguila-Arcentales S., Alvarez-Risco A., Jaramillo-Arévalo M., De-la-Cruz-Diaz M., de las Mercedes Anderson-Seminario M. (2022) Influence of social, environmental and economic sustainable development goals (SDGs) over continuation of entrepreneurship and competitiveness. *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 8, no 2, p. 73.

Dhanraj D., Vora D., Naik P. (2023) A Multiobjective Deep Learning Solution for Optimizing Cooling Rates of Urban Courtyard Blocks in Hot Arid Zones. *International Journal of Sustainable Development & Planning*, vol. 18, no 10.

Dhifaoui Z., Ncibi K., Gasmi F., Alqarni A. A. (2023) The Nexus between Climate Change and Geopolitical Risk Index in Saudi Arabia Based on the Fourier-Domain Transfer Entropy Spectrum Method. *Sustainability*, vol. 15, no 18, p. 13579.

Dong C., Xu G., Han G., Bethe, B. J., Xie W., Zhou S. (2022) Recent developments in artificial intelligence in oceanography. *Ocean-Land-Atmosphere Research*.

Dube T., Huhn A. L., Guimarães Nobre G., Moyo E. N., Enenkel M. (2024) Incorporating indigenous knowledge systems-based climate services in anticipatory action in Zimbabwe: an ex-ante assessment. *Frontiers in Climate*, no 6, p. 1301908.

Elufioye O. A., Ike C. U., Odeyemi O., Usman F. O., Mhlongo N. Z. (2024) Ai-Driven predictive analytics in agricultural supply chains: a review: assessing the benefits and challenges of ai in forecasting demand and optimizing supply in agriculture. *Computer Science & IT Research Journal*, vol. 5, no 2, pp. 473–497.

Fomin V., Putrimas M., Žižiūnas T., Laužikas R. (2022) Reference architecture for AI-based urban heritage preservation risks monitoring tool. *Baltic Journal of Modern Computing*, vol. 10, no 2, pp. 142–158.

Gómez-Villarino M. T., Briz T. (2022) With sustainable use of local inputs, urban agriculture delivers community benefits beyond food. *California Agriculture*, vol. 76, no 4.

Guermazi I., Gharbi M. W. (2024) Analyzing the achievement of the sustainable development goals (SDGs) in Saudi Arabia and the impact of the COVID-19 pandemic. *Journal of Financial Reporting and Accounting*.

Hamdan A., Ibekwe K. I., Etukudoh E. A., Umoh A. A., Ilojianya V. I. (2024) AI and machine learning in climate change research: A review of predictive models and environmental impact. *World Journal of Advanced Research and Reviews*, vol. 21, no 1, pp. 1999–2008.

Hannan M. A., Al-Shetwi A. Q., Ker P. J., Begum R. A., Mansor M., Rahman S. A., Dong Z. Y., Tiong S. K., Mahlia T. M. I., Muttaqi K. M. (2021) Impact of renewable energy utilization and artificial intelligence in achieving sustainable development goals. *Energy Reports*, no 7, pp. 5359–5373.

Hasas A., Hakimi M., Shahidzay A. K., Fazil A. W. (2024) AI for Social Good: Leveraging Artificial Intelligence for Community Development. *Journal of Community Service and Society Empowerment*, vol. 2, no 02, pp. 196–210.

Iheanetu K. J. (2022) Solar photovoltaic power forecasting: A review. *Sustainability*, vol. 14, no 24, p. 17005.

İpek Z. H., Gözüm A. I. C., Papadakis S., Kallogiannakis M. (2023) Educational Applications of the ChatGPT AI System: A Systematic Review Research. *Educational Process: International Journal*, vol. 12, no 3, pp. 26–55.

Joel O. S., Oyewole A. T., Odunaiya O. G., Soyombo O. T. (2024) Leveraging artificial intelligence for enhanced supply chain optimization: a comprehensive review of current practices and future potentials. *International Journal of Management & Entrepreneurship Research*, vol. 6, no 3, pp. 707–721.

Kannengießer S. (2023) From Millennium Development Goals to Sustainable Development Goals: Transforming development communication to sustainability communication. *Studies in Communication Sciences*, vol. 23, no 1, pp. 121–135.

Karimi H., Adibhesami M. A., Hoseinzadeh S., Salehi A., Groppi D., Garcia D. A. (2024) Harnessing Deep Learning and Reinforcement Learning Synergy as a Form of Strate-

gic Energy Optimization in Architectural Design: A Case Study in Famagusta, North Cyprus. *Buildings,* vol. 14, no 5, p. 1342.

Karpov O. E., Pitsik E. N., Kurkin S. A., Maksimenko V. A., Gusev A. V, Shusharina N. N., Hramov A. E. (2023) Analysis of publication activity and research trends in the field of ai medical applications: Network approach. *International Journal of Environmental Research and Public Health*, vol. 20, no 7, p. 5335.

Kass J. M., Takashina N., Friedman N. R., Kusumoto B., Blair M. E. (2022) Idea paper: Improving forecasts of community composition with lightweight biodiversity monitoring across ecological and anthropogenic disturbance gradients. *Ecological Research*, vol. 37, no 4, pp. 466–470.

Khateb A. M., Alkhaibari S. A. (2023) Cross-sectional investigation of mycological diagnosis challenges in Saudi Arabia. *Frontiers in Cellular and Infection Microbiology*, no 13, p. 1203892.

Kondepudi S., Kondepudi R. (2018) What constitutes a Smart City? In *E-Planning and Collaboration: Concepts, Methodologies, Tools, and Applications,* vols. 1–3.

Kumar N., Sundaram K., Reena R., Madhumathi S. (2023) Optimizing Energy Consumption in Smart Homes Using Machine Learning Techniques. *E3S Web of Conferences*, no 387, p. 02002.

Kurnia S., Alamsyahbana M. I., Chartady R., Arifin S. V., Sesaria M. I. (2023) Circular Solutions for Decent Work and Economic Growth: Lessons from Sustainable Development Goals (SDG) 8. *Academia Open*, vol. 8, no 1, pp. 10–21070.

Limberger R., Daugaard U., Choffat Y., Gupta A., Jelic M., Jyrkinen S., Krug R. M., Nohl S., Pennekamp F., van Moorsel S. J. (2024) Biodiversity increases the forecastability of species abundances in changing environments. *BioRxiv*, 2024.

Luppicini R., Walabe E. (2021) Exploring the socio-cultural aspects of e-learning delivery in Saudi Arabia. *Journal of Information, Communication and Ethics in Society*, vol. 19, no 4, pp. 560–579.

Ma H., Vo T. V., Leong T.-Y. (2023) Hierarchical Reinforcement Learning with Human-AI Collaborative Sub-Goals Optimization. *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pp. 2310–2312.

Marji N., Kohout M., Chen L., Isik G. E., Kumar A. R. (2024) AI-enabled transition to smart European cities. *Acta Polytechnica CTU Proceedings*, no 46, pp. 85–93.

Maroju R. G., Choudhari S. G., Shaikh M. K., Borkar S. K., Mendhe H. (2023) Application of Artificial Intelligence in the Management of Drinking Water: A Narrative Review. *Cureus*, vol. 15(11).

Mashhood M., Salman H., Amjad R., Nisar H. (2023) The Advantages of Using Artificial Intelligence in Urban Planning–A Review of Literature. *STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH*, vol. 5, no 2, pp. 1–12.

Matsui T., Suzuki K., Ando K., Kitai Y., Haga C., Masuhara N., Kawakubo S. (2022) A natural language processing model for supporting sustainable development goals: translating semantics, visualizing nexus, and connecting stakeholders. *Sustainability Science*, vol. 17, no 3, pp. 969–985.

Mazzetto S. (2022) Sustainable heritage conservation in the Gulf Regions. *Proceedings of International Structural Engineering and Construction*, no 9, p. 2.

Meng F., Weng K., Shallal B., Chen X., Mourshed M. (2018) Forecasting algorithms and optimization strategies for building energy management & demand response. *Proceedings*, vol. 2, no 15, p. 1133.

Milicevic N., Kalas B., Djokic N., Malcic B., Djokic I. (2024) Students' Intention toward Artificial Intelligence in the Context of Digital Transformation. *Sustainability*, vol. 16, no 9, p. 3554.

Mishra S. P., Mishra D. P., Mishra S. (2023) The etymology of gender violence (SDG-5) in anthropocene: India. *Journal of Applied Life Sciences International*, vol. 26, no 3, pp. 53–69.

Mohammadi S. O., Kalhor A. (2021) Smart fashion: A review of AI applications in virtual try-on & fashion synthesis. *Journal of Artificial Intelligence*, vol. 3, no 4, p. 284.

Molina-Molina J. C., Salhaoui M., Guerrero-González A., Arioua M. (2021) Autonomous marine robot based on AI recognition for permanent surveillance in marine protected areas. *Sensors*, vol. 21, no 8, p. 2664.

Mooney T. A., Di Iorio L., Lammers M., Lin T.-H., Nedelec S. L., Parsons M., Radford C., Urban E., Stanley J. (2020) Listening forward: approaching marine biodiversity assessments using acoustic methods. *Royal Society Open Science*, vol. 7, no 8, p. 201287.

Mraz A. L., Tumwebaze I. K., McLoughlin S. R., McCarthy M. E., Verbyla M. E., Hofstra N., Rose J. B., Murphy H. M. (2021) Why pathogens matter for meeting the united nations' sustainable development goal 6 on safely managed water and sanitation. *Water Research*, no 189, p. 116591.

Mutambik I. (2024) The Use of AI-Driven Automation to Enhance Student Learning Experiences in the KSA: An Alternative Pathway to Sustainable Education. *Sustainability*, vol. 16, no 14, p. 5970.

Nagahisarchoghaei M., Nur N., Cummins L., Nur N., Karimi M. M., Nandanwar S., Bhattacharyya S., Rahimi S. (2023) An empirical survey on explainable ai technologies: Recent trends, use-cases, and categories from technical and application perspectives. *Electronics*, vol. 12, no 5, p. 1092.

Napolitano E. V. (2023) Intelligent technologies for urban progress: exploring the role of ai and advanced telecommunications in smart city evolution. *European Conference on Advances in Databases and Information Systems*, pp. 676–683.

Nzeako G., Akinsanya M. O., Popoola O. A., Chukwurah E. G., Okeke C. D. (2024) The role of AI-Driven predictive analytics in optimizing IT industry supply chains. *International Journal of Management & Entrepreneurship Research*, vol. 6, no 5, pp. 1489–1497.

Okulich-Kazarin V., Artyukhov A., Skowron Ł., Artyukhova N., Dluhopolskyi O., Cwynar W. (2023) Sustainability of Higher Education: Study of Student Opinions about the Possibility of Replacing Teachers with AI Technologies. *Sustainability*, vol. 15, no 1, p. 55.

Olubiyi T. O., Jubril B., Sojinu O. S., Ngari R. (2022) Strengthening Gender Equality in Small Business and Achieving Sustainable Development Goals (SDGS): Comparative Analysis of Kenya and Nigeria. *Sawala: Jurnal Administrasi Negara*, vol. 10, no 2, pp. 168–186.

Orben A. (2020) Teenagers, screens and social media: a narrative review of reviews and key studies. *Social Psychiatry and Psychiatric Epidemiology*, vol. 55, no 4, pp. 407–414.

Pandey A. K., Selvakumar V., Lavanya P., Prabha S. L., Mageshwari S. U., Naidu K. B., Srivastava R. (2024) Optimizing Power Management in IoT Devices Using Machine Learning Techniques. *J. Electrical Systems,* vol. 20, no 5.

Pandey S., Twala B., Singh R., Gehlot A., Singh A., Montero E. C., Priyadarshi N. (2022) Wastewater treatment with technical intervention inclination towards smart cities. *Sustainability*, vol. 14, no 18, p. 11563.

Pigola A., da Costa P. R., Carvalho L. C., Silva L. F. da, Kniess C. T., Maccari E. A. (2021) Artificial intelligence-driven digital technologies to the implementation of the sustainable development goals: A perspective from Brazil and Portugal. *Sustainability*, vol. 13, no 24, p. 13669.

Ramadan R. H., Ramadan M. S., Alkad, I. I., Alogayell H. M., Ismail I. Y., Khairy N. (2022) Assessment of sustainable world heritage areas in Saudi Arabia based on climate change impacts on vulnerability using RS and GIS. *Sustainability*, vol. 14, no 23, p. 15831.

Rebahi Y., Hilliger B., Lowin P., Cardoso A. (2023) AI based predictive maintenance as a key enabler for circular economy: The KYKLOS 4.0 approach. *2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*, pp. 367–372.

Requena-Mullor J. M., López E., Castro A. J., Alcaraz-Segura D., Castro H., Reyes A., Cabello J. (2017) Remote-sensing based approach to forecast habitat quality under climate change scenarios. *PloS One*, vol. 12, no 3, p. e0172107.

Roa T. V., Triana W. A. (2021) Colección de mamíferos de referencia y exhibición del Museo de Ciencias Naturales Federico Carlos Lehmann Valencia, Cali, Colombia. *Mammalogy Notes*, vol. 7, no 1, p. 205.

Sakapaji S. C., Puthenkalam J. J. (2023) Harnessing AI for Climate-Resilient Agriculture: Opportunities and Challenges. *European Journal of Theoretical and Applied Sciences*, vol. 1, no 6, pp. 1144–1158.

Sen S., Saha S., Chaki S., Saha P., Dutta P. (2021) Analysis of PCA based adaboost machine learning model for predict mid-term weather forecasting. *Computational Intelligence and Machine Learning*, vol. 2, no 2, pp. 41–52.

Shah C. S., Krishnan S. (2024) ICT, gender inequality, and income inequality: a panel data analysis across countries. *Information Systems Frontiers*, vol. 26, no 2, pp. 709–727.

Surmeli A., Narla N. P., Hoeflin C. (2022) Social Enterprises and Sustainable Development Goals: How a Global Health Project Transformed Into a Social Venture–The Case of HERA App for Refugees. In *Research Anthology on Measuring and Achieving Sustainable Development Goals,* IGI Global, pp. 353–373.

Tawfik M., Shehata A. S., Hassan A. A., Kotb M. A. (2024) *Introducing Optimal Energy Hub Approach in Smart Green Ports based on Machine Learning Methodology*.

Theodorou A., Nieves J. C., Dignum V. (2022) Good AI for good: How AI strategies of the Nordic Countries address the sustainable development goals. *ArXiv Preprint ArXiv:2210.09010*.

Tuğaç Ç. (2023) Birleşmiş Milletler Sürdürülebilir Kalkınma Amaçlarının Gerçekleştirilmesinde Yapay Zeka Uygulamalarının Rolü. *Sayıştay Dergisi*, no 128, pp. 73–99.

Yadav M., Singh G. (2023) Environmental sustainability with artificial intelligence. *EPRA International Journal of Multidisciplinary Research (IJMR)*, vol. 9, no 5, pp. 213–217.

Yadav S. K. (2023) Socio-cultural Challenges on Federal Structure Nepal. *Patan Prospective Journal*, vol. 3, no 2, pp. 195–204.

Yi I. (2023) *Encyclopedia of the Social and Solidarity Economy: A Collective Work of the United Nations Inter-Agency Task Force on SSE (UNTFSSE),* Edward Elgar Publishing.

Yu Y., Li M., Liu L., Li Y., Wang J. (2019) Clinical big data and deep learning: Applications, challenges, and future outlooks. *Big Data Mining and Analytics*, vol. 2, no 4, pp. 288–305.

Yue Y., Li S., Legg P., Li F. (2021) Deep Learning-Based Security Behaviour Analysis in IoT Environments: A Survey. *Security and Communication Networks*, no 1, p. 8873195.

Zhang F., Cui X., Wang Z., Chen S., Liu Q., Liu C. (2020) A Systematic Study of AI Applications in Cybersecurity Competitions. *2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE)*, pp. 138–146.

Ziesche S., Agarwal S., Nagaraju U., Prestes E., Singha N. (2023) Role of artificial intelligence in advancing sustainable development goals in the agriculture sector. In *The ethics of artificial intelligence for the sustainable development goals,* Springer, pp. 379–397

Искусственный интеллект в продвижении умных городов Саудовской Аравии: обеспечение целей устойчивого развития в процессе ответа на социокультурные вызовы

Борхан Сепехри
Кафедра городского планирования и дизайна, Университет Тарбиат Модарес,
Адрес: шоссе Джалал Аль Ахмад Наср, почтовый ящик: 14115-111 Тегеран, Иран
E-mail: borhanspehri@gmail.com

Абдулазис И. Альмульхим
Кафедра городского и регионального планирования, Колледж архитектуры и планирования Университет имама Абдулрахмана бин Фейсала
Адрес: почтовый ящик 1982, Dammam, 31451, Саудовская Аравия
E-mail: aialmulhim@iau.edu.sa

Мохамад Анвар Адибхесами
Факультет архитектуры, Школа архитектуры и дизайна окружающей среды, Иранский университет науки и технологии
Адрес: Нармак, Тегеран, Иран
E-mail: adibhesami@hotmail.com
Саид Макареми
Факультет городского планирования и дизайна, Университет Шахид Бехешти
Адрес: площадь Шахид Шахриари, Эвин Тегеран, 1983969411 Иран
E-mail: s.makaremi@alumni.sbu.ac.ir

Фарзанех Эджази
Факультет социальных наук, Факультет гуманитарных наук, Шахед Университет
Адрес: Халид Фарх, Тегеран, 3319118651Иран
E-mail: farzanehejazi6@gmail.com

Исследование посвящено использованию искусственного интеллекта (ИИ) для решения социокультурных проблем в Саудовской Аравии и содействия достижению целей устойчивого развития (ЦУР). Рост ИИ значительно повлиял на различные области, включая науки об обществе и городе, что делает его универсальным инструментом для анализа сложных социальных явлений, понимания человеческого поведения и оптимизации городской инфраструктуры. Саудовская Аравия, сталкиваясь с уникальными социокультурными вызовами, стремится к устойчивому развитию, опираясь на концепцию «Видение 2030» и различные проекты умных городов, подчеркивая важность решения таких проблем, как гендерное равенство, сохранение культуры, рост молодежного населения, быстрая урбанизация и изменение климата.

В статье рассматриваются ИИ-приложения и модели, используемые для решения этих проблем и продвижения соответствующих ЦУР. Приложения, применяемые в сферах обработки естественного языка, компьютерного зрения и предиктивной аналитики, могут принести пользу в таких областях, как гендерное равенство, сохранение культуры, а также образование и трудоустройство молодежи. Аналогично, ИИ может оптимизировать управление водными ресурсами, потребление энергии и городское планирование в целях решения проблем быстрой урбанизации и изменения климата. Согласовав развитие технологий ИИ с целями устойчивого развития, Саудовская Аравия может раскрыть потенциал ИИ в вопросе создания устойчивых, адаптогенных и инклюзивных «умных городов», эффективно отвечающих на социокультурные вызовы. Результаты этого исследования имеют значение не только для Саудовской Аравии, и для других стран региона, но и за его пределами, подчеркивая важность интеграции ИИ для достижения целей устойчивого развития и повышения качества жизни.

Ключевые слова: ИИ; ЦУР; Социокультурные вызовы; Умные города; Саудовская Аравия; Видение Саудовской Аравии 2030; NEOM

doi: 10.17323/1728-192x-2024-4-48-79

Social Perception of Artificial Intelligence on Twitter: A Comparative Study on Global South and Global North Countries

Sadettin Demirel
Assistant Professor, Department of New Media and Communication
Üsküdar University
Address: Altunizade, Kısıklı Cd. 16-2, 34662, Üsküdar
İstanbul, Turkey
E-mail: sadettindemirel1@gmail.com

Zindan Çakıcı
Assistant Professor, Department of New Media and Communication
Üsküdar University
Address: Altunizade, Kısıklı Cd. 16-2, 34662, Üsküdar
İstanbul, Turkey
E-mail: zindan.cakici@uskudar.edu.tr

Neslihan Bulur
Research Assistant, Department of New Media and Communication
Üsküdar University
Address: Altunizade, Kısıklı Cd. 16-2, 34662, Üsküdar
İstanbul, Turkey
E-mail: neslihan.bulur@uskudar.edu.tr

This study examines the public perception of ChatGPT on Twitter, comparing approaches from Global South (India and Turkey) and Global North (United Kingdom and United States) countries. It utilizes a text-as-data approach, applying sentiment analysis, text analysis, and topic modeling techniques on tweets in English containing "ChatGPT" collected between January 1 and June 1, 2023. The VADER sentiment dictionary is used for sentiment analysis, while n-grams and BERTopic are employed for text analysis and topic modeling, respectively. Sentiment analysis reveals a predominantly positive perception of ChatGPT across both regions, with some concerns expressed regarding potential negative impacts. Text analysis and topic modeling indicate that discussions revolve around major technology companies and leading figures, with the Global North focusing more on sectoral applications and technical aspects, while the Global South evaluates ChatGPT within local language and cultural contexts. The findings demonstrate that socio-cultural differences and technological development levels between regions are reflected in the social perception of ChatGPT. The study contributes to understanding the early reception and perception of generative AI and highlights the importance of addressing privacy concerns, developing culturally compatible solutions, and examining the potential impacts of ChatGPT in various domains.

Keywords: social perception, artificial intelligence, ChatGPT, global south, global north, Twitter/X

Introduction

Artificial intelligence (AI), having undergone rapid advancements since its inception in the mid-20th century, has emerged as a transformative and multifaceted technologi-

cal paradigm. Beyond its initial applications in text generation, AI now encompasses a wide array of capabilities, including image processing, natural language comprehension, speech recognition, predictive modeling, data analytics, and robotic systems control, thereby redefining numerous scientific and practical domains (Golan, Milliken, 2024: 1; Feuerriegel et al., 2024: 111). Among these advancements, ChatGPT, an advanced language processing model developed by OpenAI, represents a significant breakthrough. Trained on extensive datasets using state-of-the-art deep learning algorithms, ChatGPT demonstrates exceptional proficiency in generating coherent, contextually relevant, and semantically rich text. Its functionalities extend to language understanding, logical reasoning, and contextual inference, enabling it to analyze user inputs and produce text that is both syntactically fluent and conceptually meaningful. Moreover, ChatGPT offers a diverse range of applications, including knowledge dissemination, personalized recommendations, complex problem-solving, and intellectual support across various interdisciplinary domains (Floridi, Chiriatti, 2020; Zheng et al., 2021; Wang et al., 2022; Bharathi Mohan et al., 2023; Abdelghani et al., 2024). Existing scholarship has engaged with the perceptions of emerging technological innovations, such as ChatGPT, within the sociocultural and economic contexts of the Global North and Global South (Yu, 2023; Gilson et al., 2023; Tarisayi, 2024; Ishida et al., 2024). Nevertheless, comparative investigations explicitly addressing how ChatGPT is perceived in these distinct geopolitical regions remain underexplored. Such comparative studies are indispensable for illuminating the nuanced roles of cultural, economic, and social determinants in shaping the adoption and societal reception of artificial intelligence systems. This study endeavors to investigate the perception of ChatGPT, a novel technological phenomenon, across the Global North and Global South. Specifically, it employs a comparative and computational analysis of Twitter (currently known as X) data, examining user-generated content to elucidate the differential perceptions of this technology in varied cultural and regional contexts.

This research commenced with a review of the existing literature to identify critical gaps in the field and to elucidate the scientific contributions of the study. The literature review was focused on examining prior studies that investigate perceptions of artificial intelligence, with a particular emphasis on emergent tools such as ChatGPT. Through this process, studies conducted within the sociocultural and economic contexts of the Global North and Global South were analyzed, uncovering a significant paucity of comparative research explicitly addressing these two geopolitical regions. Accordingly, the research questions were crafted to address these identified deficiencies. To investigate the dominant sentiment polarities and emergent thematic structures within AI-related discourse on Twitter (RQ1), as well as to discern the convergences and divergences in AI perceptions between the Global North and Global South (RQ2), this study employed a robust methodological framework integrating topic modeling, sentiment analysis, and advanced text mining techniques. The methodological approach is delineated with precision, encompassing data collection tools, analytical procedures, and the frameworks defining the study population and sampling strategies. Further-

more, the scope of the research is circumscribed across temporal, geographical, and contextual dimensions, with a critical discussion on the implications of these limitations for the study's generalizability and interpretive depth. The findings elucidate the nuanced similarities and disparities in the perceptions of ChatGPT and artificial intelligence, providing a granular understanding of the socio-cultural and contextual factors that shape these perceptions. The concluding section synthesizes the findings with existing scholarly discourses, critically situating the study's contributions within the broader literature on the societal perceptions of ChatGPT. Moreover, the implications of the findings for both theoretical advancements and practical applications are evaluated. The study concludes by offering recommendations for future research trajectories.

The Evolution of ChatGPT and Its Research Landscape

Artificial Intelligence (AI), defined as a scientific discipline that enables computational systems to perform complex tasks mirroring human cognition, is rooted in the premise that machines can emulate human-like intelligence. Its conceptual foundations were laid by Alan Turing in his seminal article Computing Machinery and Intelligence, where he proposed the Turing Test as a criterion to evaluate a machine's ability to "think" (Turing, 1950; Golan, Milliken, 2024: 1). The formal establishment of AI as an academic discipline is attributed to the Dartmouth Conference of 1956, convened by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. This event not only introduced the term artificial intelligence but also marked a paradigm shift by institutionalizing AI as a distinct scientific domain (McCarthy et al., 2006; Demirel et al., 2024a: 220). The trajectory of AI's development has been characterized by significant technological milestones. Its initial advancements were marked by the rise of expert systems during the 1980s, followed by the rapid evolution of machine learning and deep learning methodologies in the early 2000s. Generative Artificial Intelligence (Generative AI) has emerged as one of the most transformative innovations in the history of AI, reflecting the field's continual progression (Feuerriegel et al., 2024: 111). A particularly groundbreaking development within this domain was the introduction of Generative Adversarial Networks (GANs) in 2014 by Ian Goodfellow and colleagues. GANs represent a sophisticated architecture where two neural networks, a generator and a discriminator, engage in an adversarial process to synthesize novel, high-fidelity data (Goodfellow et al., 2020). Generative AI, powered by architectures such as GANs, has enabled the creation of original images, text, and other complex data types by leveraging vast training datasets. This innovation has redefined the scope of AI, automating creative processes and facilitating the generation of unprecedented content. Its applications span diverse fields, including medical imaging, artistic creation, game design, and data augmentation, significantly expanding the operational and theoretical boundaries of AI (Creswell et al., 2018). The transformative potential of generative AI positions it as a critical inflection point in the evolution of artificial intelligence, presenting an expansive frontier for both applied and theoretical research.

Artificial Intelligence (AI) has exhibited substantial utility across a wide spectrum of domains, including but not limited to finance (Lin, 2019), logistics (Woschank et al., 2020), data security (Li et al., 2020), and automation (Wang, Siau, 2019). These applications have catalyzed unprecedented advancements in operational efficiency, strategic decision-making, and innovative processes across these industries. Among its multifaceted subdomains, Natural Language Processing (NLP) emerges as a pivotal discipline, focusing on enabling computational systems to comprehend, interpret, and manipulate human language with increasing sophistication (Kumar, Kumar, 2024; Verma, 2023). A paradigmatic innovation within the realm of NLP is ChatGPT, a generative AI model developed by OpenAI. This model represents a significant milestone in the Generative Pre-trained Transformer (GPT) series. The initial iteration, GPT, was unveiled in 2018, subsequently succeeded by GPT-2, GPT-3, and GPT-4, each iteration demonstrating exponential improvements in linguistic comprehension, contextual reasoning, and generative capabilities. These advancements have been validated through extensive empirical research (Zheng et al., 2021; Wang et al., 2022; Bharathi Mohan et al., 2023; Abdelghani et al., 2024). GPT-3, distinguished by its architecture comprising 175 billion parameters, established itself as a benchmark for large-scale language modeling, attaining unprecedented accuracy across diverse NLP tasks. ChatGPT operationalizes GPT-3 within a user-centric interface, excelling in generating coherent, human-like dialogues. This achievement underscores a pivotal juncture in AI research, amplifying the utility and societal implications of advanced language models (Cai et al., 2023; Tian et al., 2024). The introduction of GPT-4 signifies a groundbreaking advancement within this lineage, incorporating multimodal capabilities to process both text and visual inputs while producing contextually relevant textual outputs. Academic inquiries into GPT-4 have highlighted its unparalleled proficiency in in-context learning, situating it as the most robust generative AI model to date (Singh et al., 2023b; Carpenter et al., 2024; Floridi, Chiriatti, 2020). These technological breakthroughs not only delineate the trajectory of AI evolution but also signify a paradigm shift in how generative models are applied across theoretical and practical contexts.

Research examining the efficacy of ChatGPT across various domains and assessing the extent to which individuals utilize this tool productively (Singh et al., 2023a; Keiper, 2023; Nedbal et al., 2023) is complemented by social science literature exploring how ChatGPT is perceived and received as a technological advancement (Oliński et al., 2024; Qin et al., 2024). These studies predominantly investigate ChatGPT's representation across diverse contexts on digital platforms such as Twitter and Reddit. The research employs methods such as sentiment analysis, textual analysis, and other computational approaches to analyze the multifaceted representations of ChatGPT. A key finding is that ChatGPT discussions often exhibit an industry-centric orientation, characterized by negative emotional valence, whereas #GPT3 discussions span a wider thematic range and reflect positive sentiment. Furthermore, technology-focused communities tend to address the technical dimensions of ChatGPT, while other communities discuss broader societal implications, such as job displacement. Prominent

discussion topics include Education, Bard, Google Search, OpenAI, Marketing, and Cybersecurity, where the emotional engagement across these subjects demonstrates a gradual transition from initially negative sentiments to more optimistic attitudes over time (Zou et al., 2023; Koonchanok et al., 2024; Qi et al., 2024). Within social science literature, studies on ChatGPT's perceived representations typically focus on specific geographic or cultural contexts, rather than systematically comparing cross-national perceptual differences.

The perceptions, attitudes, and usage patterns of ChatGPT and similar artificial intelligence tools may differ between the contexts of the Global North and the Global South. The Global North refers to developed regions, including North America, Western Europe, and Australia (Müller, 2018: 734). These regions are characterized by high living standards, significant economic productivity, and technological innovation (Leal, Harder, 2021: 5). The Global North is distinguished by its digital production and innovation capacity, playing crucial roles in international decision-making processes from economic, political, and technological perspectives. In contrast, the Global South includes developing regions such as Africa, Latin America, South Asia, and Southeast Asia (Pineda, Mishra, 2023: 865-866). These regions often face limited technological capacity, low economic productivity, and widespread social inequalities (Haelewaters et al., 2021: 1). Due to the digital divide, the Global South benefits less from technological advancements in comparison to the Global North. These regions primarily focus on development-oriented policies. In this framework, the Global North remains central, while the Global South remains marginalized in environmental and political aspects (Molosi-France, Makoni, 2020:10-11; Haswell et al., 2024: 2). A review of the literature indicates significant differences in technology-related perceptions, attitudes, and usage between the countries of the Global North and the Global South (Archibugi, Pietrobelli, 2003; Schembri, Petit, 2009; Koskinen et al., 2019; Cai, 2022; Gordon, 2022; Mudaly, Chirikure, 2023; Boyles et al., 2024; Flacher et al., 2024; Kumar, 2024). The primary reason for these disparities lies in the historical and ongoing social, cultural, economic, and political exploitation of the Global South by the Global North (Mudaly, Chirikure, 2023; Portes Virginio et al., 2023; Gräbner-Radkowitsch, Strunk, 2023; Sultana, 2023; Sovacool, 2023; Fabian, 2023).

As evidenced, perceptions, attitudes, and usage patterns concerning technology differ between the contexts of the Global North and the Global South. In the current literature, studies focusing on the perception of new artificial intelligence systems, such as ChatGPT, are primarily concentrated in Global North countries, including the United States, Germany, Australia, and the United Kingdom (Yu, 2023; Gilson et al., 2023; Kasneci et al., 2023; Koonchanok et al., 2023; Klinger et al., 2023). In the context of the Global South, such research remains limited and has mostly been conducted in recent times (Aktay et al., 2023; Gondwe, 2023; Tarisayi, 2024; Ishida et al., 2024). Within the literature addressing societal perceptions of AI tools like ChatGPT, there is a conspicuous lack of comparative studies examining Global North and Global South countries. Such comparative research is crucial to uncover the influence of regional cultural, economic,

and social factors on the adoption and perception of AI systems. This research aims to systematically investigate the societal perceptions of ChatGPT across the Global North and Global South by utilizing text mining techniques, including sentiment analysis, text analysis, and topic modeling, on relevant Twitter interactions. Accordingly, by analyzing AI-related posts about ChatGPT made by Twitter users from both the Global North and the Global South, this study seeks to conduct a comparative examination of how this new technology is perceived in different cultural and regional contexts.

Materials and Method

To examine how generative artificial intelligence is received within the Twitter ecosystem, the text-as-data approach (Grimmer, Stewart, 2013) was preferred. Within the framework of the text-as-data approach, text is considered as data, and computational measurements and calculations are performed on the text to reveal prominent words and themes (Benoit, 2020; Grimmer, Stewart, 2013). In these processes, sentiment analysis (Zhang, Liu, 2017), text analysis (Welbers et al., 2017), and topic modeling (Günther, Quandt, 2016) techniques, which are frequently used in text mining, were employed.

The study has two main research questions. The first question focuses on which type of sentiment and topics the concept of generative artificial intelligence is associated within the Twitter. The second research question interrogates the similarities and differences between the perceptions of AI technologies among Twitter users from the Global North and Global South

RQ1 What is the prevalent sentiment and associated topics in posts related to artificial intelligence on Twitter.

RQ2) What are the differences and similarities in the perceptions of generative artificial intelligence among Twitter users from Global North and Global South countries?

VADER Sentiment Analysis

Sentiment analysis, also known as opinion mining, is simply a text classification process (Liu, 2020; Zhang, Liu, 2017). This classification process is performed on various texts, from social media to the internet, documents, and speeches (Pang, Lee, 2008; Pozzi et al., 2017; Zhao et al., 2016). With sentiment analysis, texts are generally classified into two categories (positive, negative) or three categories (positive, negative, and neutral) (Liu, 2020). This process varies according to the technique used, the text, and the purpose of the application. In this study, the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment dictionary, developed for conducting sentiment analysis on social media posts, was employed (Hutto, Gilbert, 2014). VADER is a lexicon-based sentiment dictionary. It can provide more accurate results than human coders on social media texts in the classification tasks (Hutto, Gilbert, 2014; Ribeiro et al., 2016). It provides classification of text from social media or web environments into positive, negative, and neutral categories (Hutto, Gilbert, 2014).

VADER is frequently used in many studies in terms of detecting polarity and classifying texts to sentiment categories, from political communication to public opinion research, health communication to crisis communication. Demirel et al. (2024b) used VADER to compute polarity score and explore political discourse of Twitter users after Turkish government converted the Hagia Sophia Museum to mosque. Similarly, Nisch (2024) utilized VADER to analyze public discussions and debates regarding Finland's decision to join NATO. Moreover, Dahal, Kumar & Li (2019) instrumentalized VADER to reveal overall sentiment and attitudes in climate change related tweets whereas VADER were also applied on text data to detect health misinformation (Zhong, 2023), public health concerns (Zolnoori, et al., 2021), public perceptions during and after the disasters (Li et al., 2023; Zhai et al., 2024) and the social media perceptions of COVID-19 vaccines (Kahraman et al.,2023; Saleh et al., 2023). In line with this study, VADER is also used to measure social media perception and receptions of novel digital trends such as concept of metaverse (Gündüz, Demirel 2023), public reactions regarding ChatGPT's impact on education (Fütterer et al., 2023), public perception of electric cars on Reddit (Ruan, Lv, 2022), public discussions on internet of things (Zubiaga et al., 2018). As it was seen that VADER dictionary was widely employed to measure public opinion and perception, there is a limited number of studies that use VADER's sentiment classification to compare perception of artificial intelligence in particularly Global South and Global North. Hence, to fill this gap in the literature, VADER sentiment analysis was performed on all tweets and subsequently calculated based on certain countries (India, Turkey, UK, USA) and region (Global South, Global North). Consequently, sentiment analysis allowed for the acquisition of Twitter users' opinions and approach regarding generative artificial intelligence technology, specifically in the context of ChatGPT, the leading chatbot in the field.

Text Analysis and Topic Modelling

In the text analysis process, frequently used terms in tweets were obtained to examine the social perception produced about the ChatGPT tool and generative artificial intelligence technology. In text analysis, tweets were first passed through a tokenization step and divided into words (Hvitfeldt, Silge, 2022; Jo, 2019; Silge, Robinson, 2017; Wiedemann, 2016). Unnecessary terms among the obtained words were cleaned using a stopwords list (Benoit et al., 2020). Subsequently, using the n-gram technique, which is frequently employed in text mining, consecutive paired words called bigrams were identified, and the word frequencies of these words were calculated (Silge, Robinson, 2017; Wiedeman, 2016; Welbers et al., 2017). R software (R Core Team, 2023) and the Quanteda package (Benoit et al., 2018) were used for the calculations.

Topic modeling, on the other hand, is an unsupervised text classification technique (Grimmer, Stewart, 2013; Günther, Quantd, 2016; Silge, Robinson, 2017). According to the principle of topic modeling, each text or document can consist of multiple topics, and these topics (themes, clusters, etc.) can be obtained through the frequent and joint

use of words within the text (DiMaggio et al., 2013; Ramage et al., 2009). In this way, the text is reduced to specific categories and made meaningful. In the literature, the topic modeling technique is frequently used on newspaper headlines, tweets, and documents (DiMaggio et al., 2013; Jónsson, Stolee, 2015; Lucas et al., 2015; Quinn et al., 2010; Rabitz et al., 2021; Yin et al., 2022). Within the scope of this study, BERTopic (Grootendorst, 2022) was utilized for the topic modeling process using Python. Unlike other topic modeling approaches, BERTopic is a transformer-based topic modeling technique that forms the basis of technologies like ChatGPT. It relies on machine learning technique for natural language processing called BERT (Bidirectional Encoder Representations from Transformers). In this study, zero-shot topic modeling was preferred via BERTopic. In zero-shot topic modeling, BERTopic classifies words related to predefined categories by the researcher according to their usage within the text (Grootendorst, 2022). In this study, topic modeling was performed on 20 predefined topics within the tweets. The findings of the previously conducted text analysis and basic-level topic modeling were utilized to discover the 20 categories or topics. Subsequently, using the zero-shot topic modeling process, 20 topics with high frequency and the words within these topics were determined within the tweet data.

Data Collection & Pre-processing

To understand public perception of artificial intelligence within the context of ChatGPT, more than half a million tweets (n = 750,000) containing "ChatGPT" were initially collected. The data collection process was carried out between January 1, 2023, and June 1, 2023, using Twitter's API (Application Programming Interface) service (Barrie, Ho, 2021). The presence of more accurate software for text processing and analysis in English compared to other languages influenced the selection of only English tweets within the scope of the study. To identify the emerging social perceptions about ChatGPT, data collection began one month after ChatGPT-3 was launched by OpenAI on January 1th 2023 and it is ended on June 1st 2023. The main factors that limit the time frame is the Elon Musk's buyout of Twitter (Paul, Milmo, 2022) and which results in subsequently discontinuation of Academic API for researchers (Calma, 2023).

Within the scope of the study, all English tweets were collected during the mentioned time interval. To focus directly on users' original ideas, retweeted tweets were removed from the data. Additionally, in line with the research objective, only tweets shared from specific regions and countries were filtered. The United States and the United Kingdom were selected from the Global North, while India and Turkey were chosen from the Global South. The number of Twitter users and amount of time Twitter is used in these countries (*X/Twitter*, 2024) were the main reasons that these countries were selected. Using the *user location* variable available in the dataset obtained from the Twitter API, tweets shared from selected countries were filtered for the analysis. In the end, total of 528,137 tweets were included in the study. The distribution of tweets by country is shown in Table 1.

Table 1. Number of tweets by country

Country	Number of tweets	% of tweets
India	76793	14.54
Turkey	27610	5.22
UK	69372	13.13
USA	354362	67.09

In another step of the pre-processing stage, common words (the, you, an, a, etc.), links, and punctuations were removed from the text data before text analysis by using a list of stop words in English (Benoit et al., 2018; Günther, Quandt, 2016). This allowed for easy identification of prominent terms in the tweets.

Limitations of the Study

This study has several limitations. First, It only focused on single tweets in English and with proper user location data due to the lack accurate and consistent text mining solutions such as VADER in other language. Secondly, only Twitter social platform was used to scrutinize public perception of AI despite there are several social networks (e.g. Instagram, Telegram, Reddit, etc.) provides users with avenues to express their opinion. For these reasons, we are aware that our findings only reflect only Twitter users from the selected countries within the first five months of the ChatGPT's launch. Therefore, with the widespread use of these technologies in time, public perceptions and opinions might change. Nevertheless, computational analysis of AI perception of Twitter users in the first five months also provide valuable insights to further delve into perception AI and acceptance of this novel technology between the countries and regions. More importantly despite certain limitations, this study findings may enable comprehensive theory building to understand social perception of artificial intelligence. To remedy the representation issue and dataset imbalance, we also analyzed tweets data per country in sentiment and text analysis. Moreover, social media users in Twitter are highly vocal and active when it comes to the new trends and discussions. Thus, Twitter features that provide digital environment to discuss new and novel subjects such as, A. I. somewhat minimizes the limitation of this study. This makes the Twitter desirable sphere to carry out of a study that focuses on user-generated contents and users' perception of novel technologies and trends.

Results

The findings were presented in three sections. First one focuses on polarity of tweets, while the second concerns with textual features of ChatGPT related tweets from both regions. Last section of findings shows the main topics related with artificial intelligence on Twitter.

VADER Sentiment Analysis

The sentiment analysis results indicate that tweets predominantly exhibit a positive sentiment. Specifically, approximately 50% of the tweets are classified as having positive polarity, while 32% are neutral, and 18.8% are negative. Given that "ChatGPT" is a central keyword in all tweets analyzed, the data presented in Table 2 highlight a general inclination among Twitter users towards this tool and, more broadly, the concept of generative artificial intelligence. These findings suggest that, within the temporal scope of the study, Twitter users tend to hold a favorable perception of ChatGPT.

Table 2. Number of tweets by VADER sentiment categories

Sentiment	Number of tweets	% of tweets
Positive	255120	49.12
Neutral	166159	31.99
Negative	98048	18.87

The following tweets can be given as examples of the sentiment categories in Table 2:

Negative

"ChatGPT's 'liberal' bias allows hate speech towards GOP, men: research."

"ChatGPT replaces humans in 50% of US companies that use the AI bot: Report\ nNow this is a disturbing news."

Positive

"Are you a wedding planner? Use ChatGPT to create beautiful portfolios, marketing materials, and social media posts to attract more clients!"

"Yes, ChatGPT is impressive. This is not news. But I just prompted ChatGPT to write me story about a bucket winning an academy award... and it is so good I can't keep it to myself."

According to the distribution of tweets' average sentiment categories by country in Table 3, the proportion of positive tweets is higher than the proportion of tweets in other categories in all countries. Turkey has the highest percentage of positive tweets at 51.10%, while India has the lowest at 46.01%. In terms of negative tweet percentages, the UK ranks first with 19.47%, while India and the USA have the lowest negative tweet percentages at 18.80% and 18.75%, respectively.

Table 3. Number of tweets by country and VADER sentiment category

Ülke	Mean Sentiment	Negative	Neutral	Positive
India	0.160	18.80	35.17	46.01
Turkey	0.191	19.38	29.51	51.10
UK	0.181	19.47	30.13	50.38
USA	0.175	18.75	31.80	49.44

Table 3 shows that the difference in the distribution of tweets' sentiment categories between global north and global south countries consists of single-digit numbers. In the light of this findings, it can be said that users from the four countries have similar sentiment states towards artificial intelligence, particularly ChatGPT.

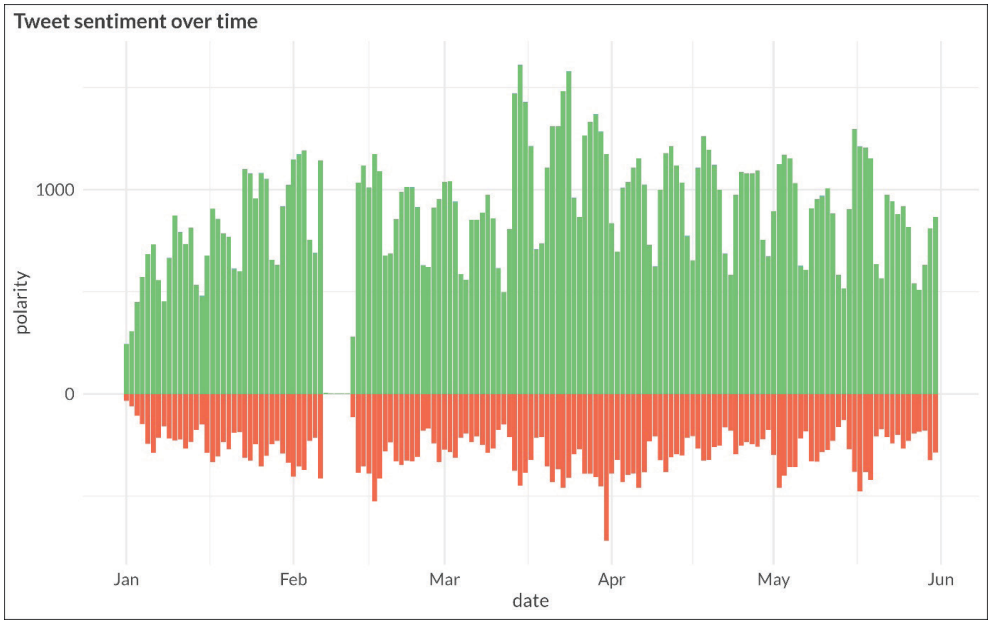


Figure 1. Tweet sentiment over time (green = positive polarity, red = negative polarity)

In line with the findings in Table 2 and Table 3, the polarity values of tweets over time in Figure 1 also indicate that positive tweets are in the majority. Although certain fluctuations are observed in the polarity values over time, a positive polarity is observed in most of the tweets shared about ChatGPT during the first five months of 2023.

In Figure 2, the polarity changes in the tweets of users in India, the UK, and the USA show similarities. On the other hand, a brief increase in negative tweets was observed

in early April among users in Turkey. The differences between countries have also been influenced by the quantitative variation in tweets.

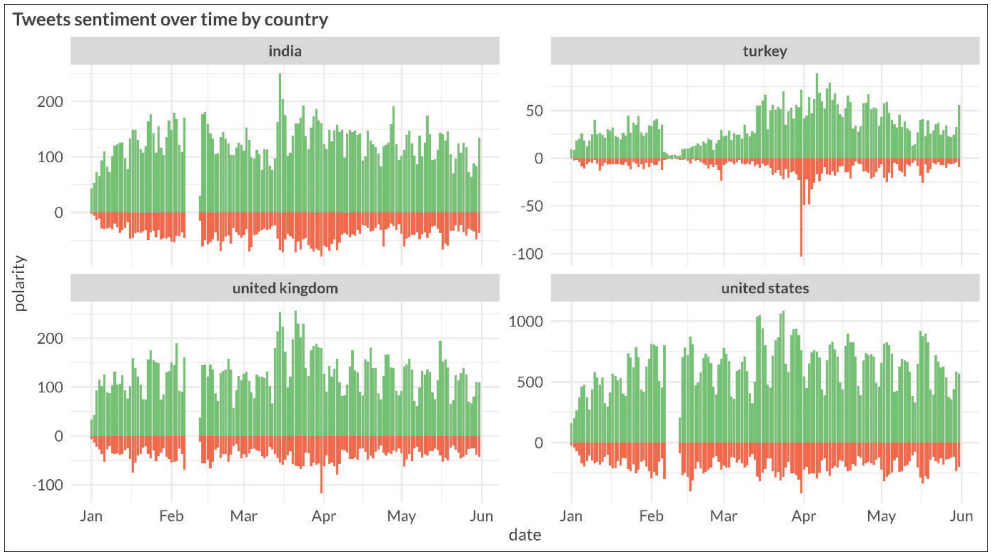


Figure 2. Tweet sentiment over time by country (green = positive polarity, red = negative polarity)

The VADER dictionary can not only provide polarity scores but also classify tweets into sentiment categories such as positive and negative. The prominent terms (words and hashtags) in tweets of both polarities can reveal the public perception of generative artificial intelligence technology constructed within the minds of Twitter users.

Figure 3 shows the prominent unigram (single word) words in tweets belonging to the positive and negative sentiment categories, indicating the words associated with the concept of artificial intelligence. In negative tweets, words like *banned, fake, wrong, threat, cheating, dangers, fraud, concern, privacy* represent the concerns and fears about the emerging new technology. Considering the effects of ChatGPT on educational activities, the business world, and the knowledge ecosystem, these reactions are not unfounded. On the other hand, there is a hype and sympathy towards generative artificial intelligence technologies, particularly ChatGPT. Looking at the words with positive connotations such as *create, business, love, powerful, productivity, boost, amazing, perfect, enhance*, an optimistic trend is also observed.

In line with Figure 3, the high-frequency hashtags used in tweets of different polarities can provide better results regarding the content of the texts. In Figure 4, the frequently used hashtags in negative tweets are related to topics such as *cybersecurity, cybercrime, information security, privacy, data security, phishing, malware, misinformation, fakenews, hackers, and cheating.*

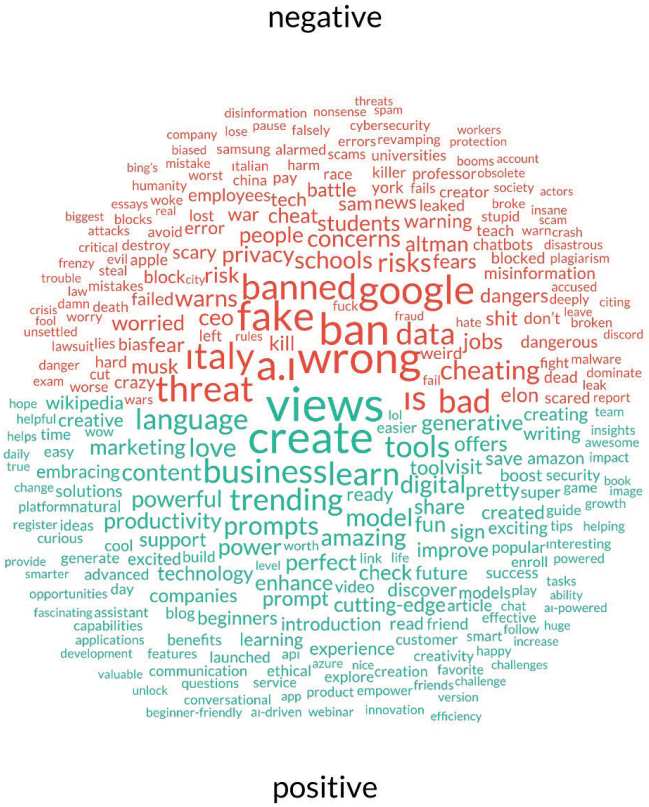


Figure 3. Frequently used word by sentiment (green = positive polarity, red = negative polarity)

As with the frequently used words in tweets (Figure 3), the problems that artificial intelligence may cause in the news industry, digital privacy, and cybersecurity have been frequently expressed in tweets with related hashtags. On the other hand, in the positive tweets of Twitter users, hashtags focusing on the innovations and benefits that ChatGPT will bring to the *business world, marketing, and SEO* activities have been used. Additionally, topics such as *digital transformation, productivity, coding, and content production* have been frequently discussed on Twitter.

According to the sentiment analysis results performed on the Twitter posts collected within the framework of the study's sample, there is a positive perception of the artificial intelligence language model ChatGPT. In fact, this trend shows the presence of an optimism bordering on hype regarding ChatGPT and related generative artificial intelligence technologies, regardless of the country where users are located. On the other hand, despite the dominance of positive tweets in the distribution of sentiment categories, it is observed that the potential problems and negativities that this technology may cause are also expressed in tweets.

Figure 4. Frequently used hashtags by sentiment (green = positive polarity, red = negative polarity)

Within 6 months after its launch, ChatGPT is positively approached and accepted by users, and opinions are shared that it will positively impact the business world, marketing, advertising, content production, SEO, and similar fields. On the other hand, concerns such as education (e.g. cheating in exams), cybersecurity issues, digital privacy, fake news production, etc., which this technology may create, also stand out. The prominent words and hashtags on both sides demonstrate the social perceptions constructed about ChatGPT. However, to be able to examine these perceptions thoroughly for each country, employing computational text analysis on tweets is a must. Latter section present frequency analysis of top lexical features, which can offer better context regarding the public perception of artificial intelligence in these countries.

Text Analysis

Word Frequency Analysis

Figure 5 shows the results of the text analysis of tweets related to ChatGPT for four different countries and 2 regions: India, Turkey (Global South), United Kingdom, and United States of America (Global North).

In both Global South (Turkey and India) and Global North (United Kingdom and USA) countries, the terms "elon_musk", "sam_altman", and "search_engine" are observed with high frequencies. This situation indicates that generative artificial intelligence technologies are discussed globally around similar actors and applications. Particularly, leading figures in the technology sector and search engine technologies are at the center of generative artificial intelligence discussions, regardless of geographical region.

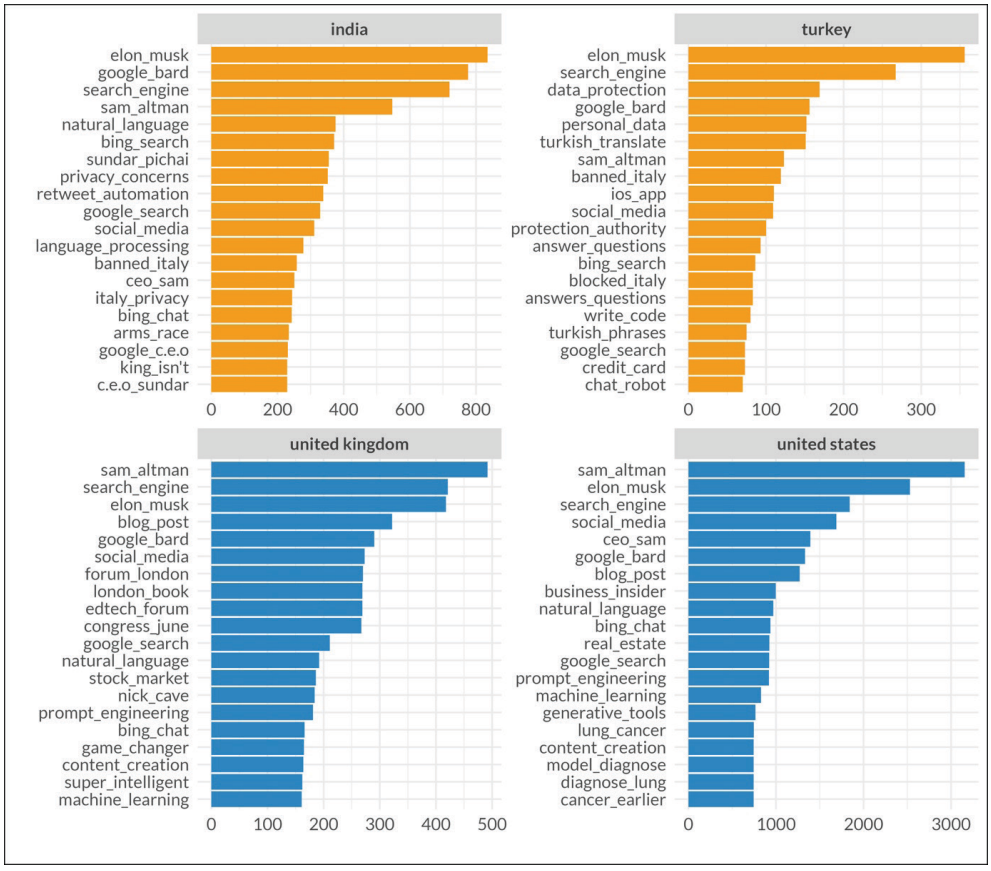


Figure 5. Top 20 bigram words by region (Blue = Global North, Orange = Global South)

The prominence of rival AI models or applications such as "Google Bard (currently named Gemini)", "Bing chat", and "Bing search" (currently known as Microsoft Copilot) in the discussions about ChatGPT in countries from both regions demonstrates a strong interest in the artificial intelligence initiatives of techno-political entities like Google and Microsoft. Therefore, with the launch of ChatGPT, the more frequent appearance of alternatives in Twitter posts indicates that technological competition has become more

diverse. On the other hand, the fact that all the mentioned organizations and actors originate from the Global North makes two main interpretations possible. The first is the technological divide between regions, and the other is that the social perceptions in the field of generative artificial intelligence bear socio-cultural traces of Global North countries.

The high frequencies of bigram words such as "data protection" and "personal data" in Turkey, and "privacy concern" in India, show that privacy concerns about generative artificial intelligence technologies are at the forefront in the Global South. The relatively less emphasis on such terms in the Global North countries suggests that privacy issues either cause less concern or the discussions focus on different dimensions in these regions. Moreover, the prominence of terms like "Turkish translate" and "Turkish phrases" in Turkey indicates that generative artificial intelligence technologies are evaluated in the context of local language and culture in the Global South. This points to the importance given not only to the global impact of technology but also to its local adaptation.

In the Global North countries, particularly in the USA, the use of words related to specific application areas such as "business insider", "real estate", and "lung cancer" shows that more detailed and sector-specific discussions are being held about the potential use cases of generative artificial intelligence. In the Global South, more general terms about the features of artificial intelligence and news about countries' ChatGPT policies are at the forefront, demonstrating how the technological development gap between regions is reflected in the discussions. While one side focuses on the features and threats of this technology, the other side discusses the sectoral integration of a newly launched tool.

The more frequent appearance of technical terms such as "prompt_engineering" and "machine_learning" in the Global North countries indicates that generative artificial intelligence technologies are addressed from a more technical perspective in these regions. It can be said that this situation is related to technological edge that Global North has in research, development and innovation.

Hashtag Frequency Analysis

In addition to words, hashtags used on Twitter to make a topic trending or reach more people can also provide rich findings for interpreting the shares about ChatGPT.

As with word frequencies, the hashtags "technology" and "tech" are observed with high frequencies in all countries. This is an expected finding because the topic of generative artificial intelligence is directly related to the general technology context. However, in this graph, specific terms such as "machinelearning" and "generativeai" stand out more prominently in the Global North. This supports the previous observation that discussions in the Global North are more technical and specific.

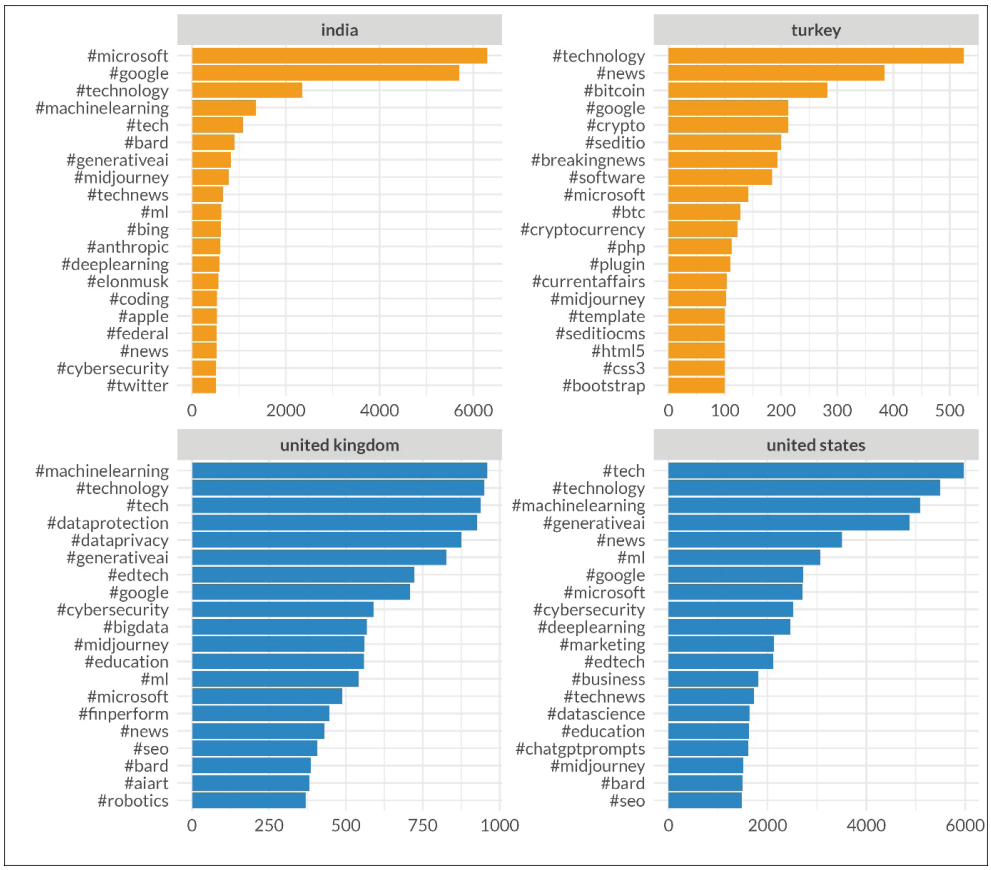


Figure 6. Top 20 hashtags by region (Blue = Global North, Orange = Global South)

In Figure 6, company names have replaced person-focused terms such as "Elon Musk" and "Sam Altman" that we saw in high-frequency words. Hashtags such as "Microsoft", "Google", and "Apple" are frequently used by Twitter users in both regions. This shows that generative artificial intelligence discussions are shaped around large technology companies. Moreover, the fact that organizations originating from the Global North are on the agenda in artificial intelligence discussions supports the previously mentioned finding about the technological infrastructure and development gap between regions.

When it comes to hashtags, privacy and data security issues emerge more prominently in the United Kingdom (#dataprotection, #dataprivacy). This indicates that the issue of privacy is not unique to the Global South and is also an important topic of discussion in the Global North. The high frequencies of the hashtags #bitcoin, #crypto, and #cryptocurrency in Turkey reveal a dimension we did not see in Figure 5. This

shows that generative artificial intelligence technologies are associated with financial technologies in the Global South. The presence of the hashtags #news or #breakingnews in all countries indicates that the topic of generative artificial intelligence is discussed in the context of current events. The fact that artificial intelligence finds a place in the news alongside user shares shows that the social perception of this concept is also shaped by the media.

The hashtags such as #education, #edtech, #seo, and #business in the Global North countries shows the interest in the potential applications of generative artificial intelligence in education, SEO, and other business areas. This supports the interest in sectoral applications mentioned in the word frequency analysis.

The presence of hashtags related to web development such as #php, #html5, and #css3 in Turkey shows that generative artificial intelligence is discussed in the context of web technologies. This is a finding that we did not see in our analysis in Figure 5, suggesting that more practical applications of technology are focused on in the Global South.

Topic Modelling

Figure 7 shows the prominent topics or themes in tweets related to ChatGPT after the topic modeling process. These themes were obtained after text analysis and basic topic modeling performed on the tweets, and subsequently, using BERTopic, the frequencies of these topics within the tweet data were calculated on a regional basis. As with the text analysis findings, Figure 7 also shows similarities between the two regions in certain topics. Topics such as machine learning, developments regarding the regulation (ban & block) of ChatGPT, prompt engineering, Google and Microsoft, coding and programming, etc., constitute the main themes of user tweets in both regions.

The ranking of topics in the findings of Figure 7 is also significant. For example, while the topic of business & innovation ranks seventh in the Global North countries, it ranks fourteenth in the Global South countries. In the topic frequencies, a significant difference between the regions is that topics related to the impact of ChatGPT on local languages and issues have also come to the forefront in the discussions in the Global South countries. For instance, the potential impact of ChatGPT on Bollywood in India, its performance in Hindi dialects, similarly, its performance in the Turkish language in Turkey, and its relationship with crypto and Bitcoin are observed as different topics. Another prominent theme is the developments in the field of artificial intelligence in China. Although we failed to reach these findings in the text analysis, it is seen that Twitter users in the Global South do not focus solely on Global North-centric technology initiatives and companies when it comes to artificial intelligence. On the other hand, there is a similarity of topics on a regional basis in tweets related to ChatGPT. These findings also align with the text analysis findings (Figure 5 and Figure 6).

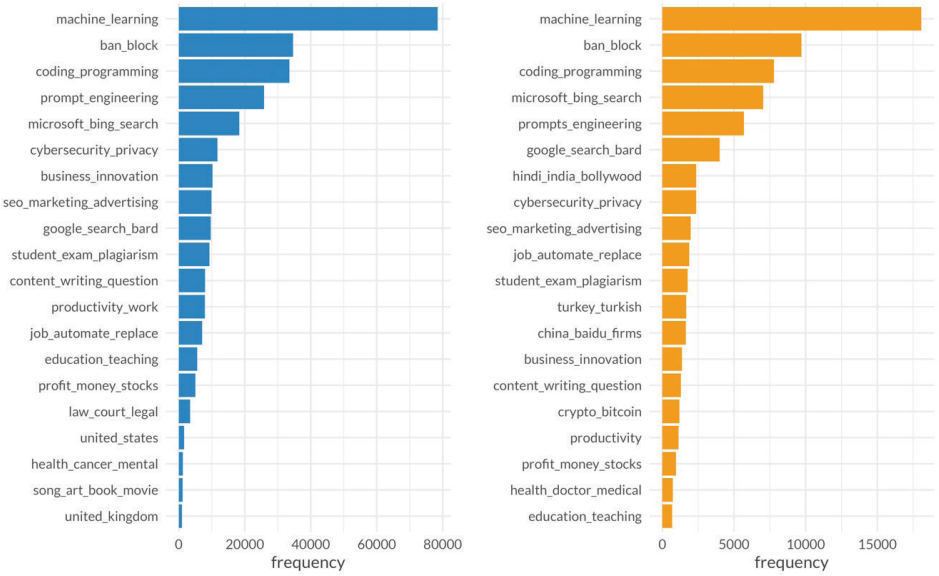


Figure 7. Topic frequency by region (blue = Global North, orange = Global South)

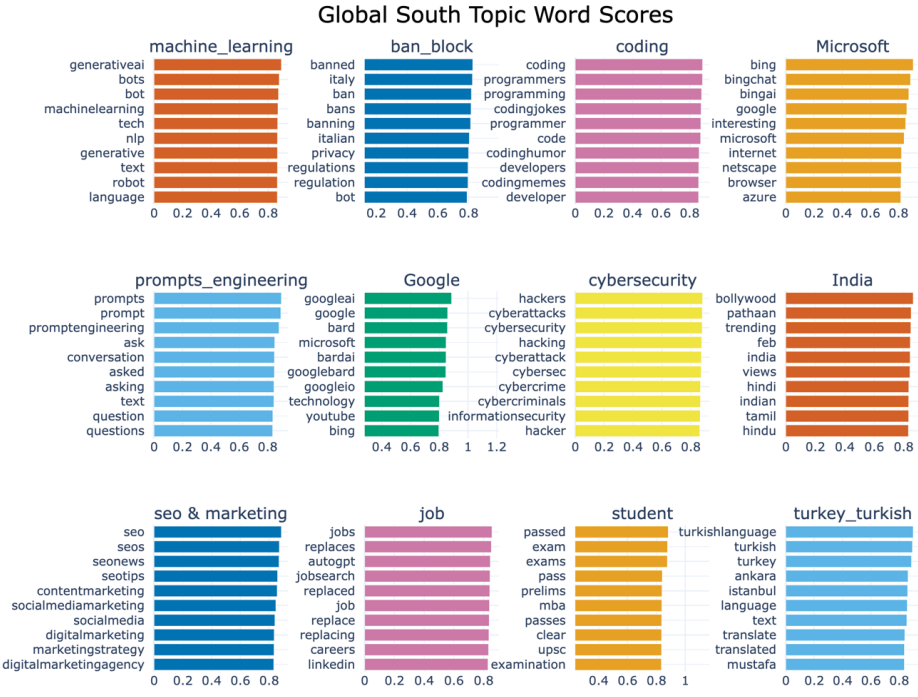


Figure 8. Top 12 Topic word scores on tweets from Global South

Figure 8 and Figure 9 show the topics discovered in tweets shared from Global South and Global North countries, respectively, and the high-frequency words within these topics. There is a high degree of similarity in both the topics and the words that constitute these topics. To better visualize the words, the 12 topics with the highest frequency were selected. Additionally, as seen in Figure 7 and Figure 8, topics such as "turkey_turkish" and "india" discussed in the Global South countries differ. The words in Figures 8 and 9 are simplified versions of the topics in Figure 7. For example, the student topic corresponds to the student_exam_plagiarism topic in Figure 7. As can be understood from the graphs, the use of ChatGPT in assignments and exams is frequently discussed in tweets in both regions. Although the topic being discussed is the same, its presence in tweets with different words also shows the different impact of ChatGPT on educational activities. In the Global South countries, terms such as "passed", "exam", and "prelims" stand out, while in the Global North countries, words like "plagiarism", "cheating", and "professor" are emphasized more. Therefore, it can be said that in the Global North countries, more focus is placed on plagiarism and academic ethics. In the coding topic, terms such as "codingjokes", "codinghumour", and "codingmemes" are frequently used in the Global South countries, while users in the Global North prefer words like "Python", "github", and "developer" in the same category. Thus, there is a difference between regions in discussions related to ChatGPT and coding. A similar situation is observed in ChatGPT discussions around the business theme. Users in the Global South view artificial intelligence as a threat to their jobs with words like "replace", "replaced", and "jobsearch", while tweets shared from the Global North countries emphasize the productivity aspect of ChatGPT more with words like "productivity", "workload", "improve", and "easier". Therefore, although similar topics are discussed in both regions, especially regarding ChatGPT, users approach the same topic from different perspectives due to their socio-cultural differences. This situation shows that when it comes to an unknown technology, the constructed social perceptions exhibit similarities as well as differences between regions and cultures.

Discussions & Conclusions

This study aimed to examine the public perception of ChatGPT on Twitter by analyzing tweets shared from Global South (India and Turkey) and Global North (United Kingdom and United States) countries by employing sentiment analysis, text analysis, and topic modeling methods.

Sentiment analysis revealed that the overall perception of ChatGPT is not negative. Given the proportion of positive posts, users in both regions displayed an optimistic attitude and some may call it "hype" towards generative AI technologies. These findings support the results of previous studies, demonstrating that artificial intelligence is largely accepted by society and generates positive expectations (Gilson et al., 2023; Kasneci et al., 2023; Yu, 2023). However, alongside the positive perceptions, concerns regarding the potential negative impacts of ChatGPT were also expressed. Notably,

apprehensions about data privacy, cybersecurity, and the generation of fake news were prominent. This indicates that the societal effects of generative AI are multidimensional.

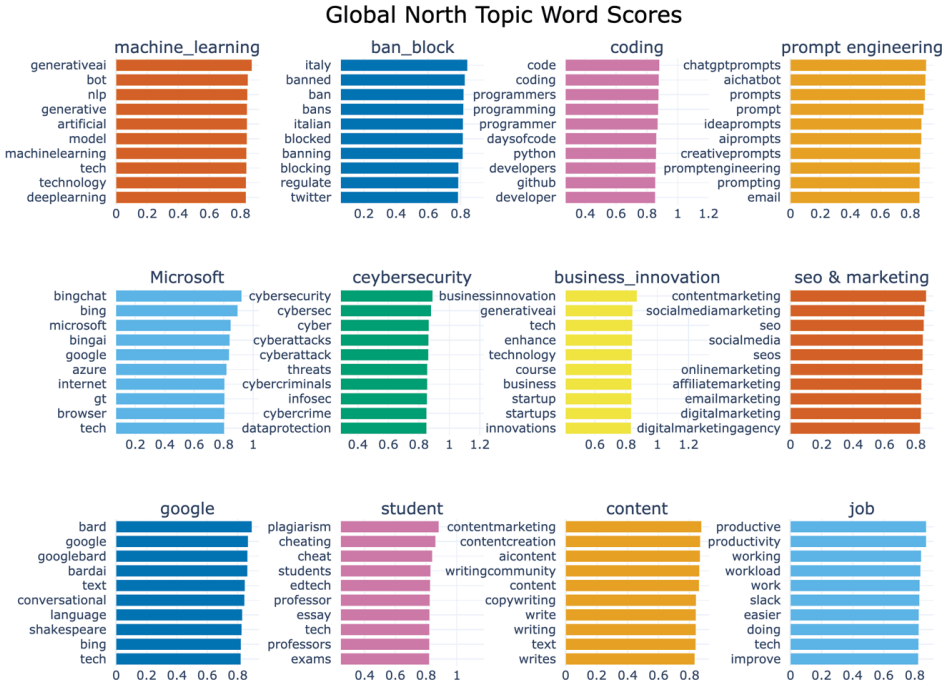


Figure 9. Top 12 Topic word scores on tweets from Global North

Text analysis and topic modeling on tweets show that ChatGPT discussions in both regions were shaped around major technology companies and leading technology figures. This finding aligns with the results of previous studies, indicating that AI discussions are driven by specific actors (Koonchanok et al., 2024; Zou et al., 2023). Furthermore, while the sectoral applications and technical aspects of ChatGPT were more frequently discussed in Global North countries, it was observed that ChatGPT was evaluated within the context of local language and culture in Global South countries. This situation demonstrates that socio-cultural differences and the level of technological development between regions are reflected in social perceptions.

One of possible implications arising from the findings is the influence of the technological divide between the Global North and Global South on the social perception of generative AI. The study reveals that discussions in the Global North focus more on the technical uses and sectoral integrations of AI, while the Global South evaluates AI within local language and cultural contexts. This suggests that the existing technological infrastructure and development disparities between regions shape how AI is perceived and

represented in social discourse. Countries with more advanced AI ecosystems and greater access to AI technologies are likely to have more informed and nuanced discussions about the capabilities and potential applications of generative AI. Conversely, countries with limited exposure to AI may focus more on the basic features and potential threats of the technology. Addressing this technological divide through increased investment, knowledge sharing, and capacity building in the Global South could help bridge the gap in AI perceptions and foster more balanced global discussions.

Another implication relates to the socio-cultural differences in AI perception between the two regions. The study finds that while similar topics are discussed in both regions, users approach them from different perspectives due to their socio-cultural backgrounds. For example, in the context of education, the Global South focuses on the impact of ChatGPT on exams and passing grades, while the Global North emphasizes academic ethics and plagiarism concerns. Similarly, regarding the impact on jobs, the Global South views AI as a threat to employment, while the Global North highlights its potential for productivity improvements. These findings underscore the importance of considering cultural contexts when designing, deploying, and governing AI systems. Developers and policymakers must be mindful of the diverse socio-cultural factors that shape AI perceptions and ensure that AI technologies are culturally sensitive and aligned with local values and priorities. Failure to do so may result in resistance, mistrust, and unequal distribution of AI benefits across regions.

The study also reveals implications for the economic dimensions of AI adoption and development. The prominence of Global North-based technology companies and initiatives in AI discussions across both regions indicates the concentration of AI power and influence in the Global North. This raises concerns about the potential for AI to exacerbate existing economic inequalities between the two regions. If the Global South remains primarily a consumer of AI technologies developed in the Global North, it may miss out on the economic benefits of AI innovation and become increasingly dependent on foreign AI providers. To counter this, countries in the Global South must invest in developing their own AI capabilities, fostering local AI ecosystems, and promoting indigenous AI research and development. This will enable them to harness the economic potential of AI, create new job opportunities, and ensure a more equitable distribution of AI benefits.

One of the strengths of this study is its examination of the social perception on AI in the both Global North and Global South countries. This approach allowed for the identification of similarities and differences between the regions. Moreover, the methods employed in the study (sentiment analysis, text analysis, topic modeling) enabled a multidimensional investigation of ChatGPT's social perception.

However, the study also has certain limitations as it is elaborated in the end of method section. Firstly, although Twitter data represents a broad user base, it may not fully reflect the entire population of a region and country. Therefore, the findings may be valid only for the Twitter users that reside in these countries or regions. Additionally, the study covers a specific period and does not examine how the social perception of ChatGPT changes over time.

In conclusion, by examining the public perception of ChatGPT in the context of Global North and Global South countries, this study has provided important findings regarding the societal perception of generative AI technologies. Future research should investigate the changes in the perceptions and receptions over time, their reflections on different social media platforms, and the factors influencing the societal acceptance of AI technologies, which will make significant contributions to this field.

References

Abdelghani R., Wang Y. H., Yuan X. et al. (2024) GPT-3-driven pedagogical agents to train children's curious question-asking skills. *International Journal of Artificial Intelligence in Education*, vol. 34, no 3, pp. 483–518.

Aktay S., Gok S., Uzunoglu D. (2023) ChatGPT in Education. *TAY Journal*, vol. 7, no 2, pp. 378-406.

Archibugi D., Pietrobelli C. (2003) The globalisation of technology and its implications for developing countries: Windows of opportunity or further burden? *Technological forecasting and social change*, vol. 70, no 9, pp. 861-883.

Baptista J., Costa D., Gonçalves S. P. (2023) Social Representations of Trans People in the Workplace. *Social Sciences*, vol. 12, no 8, p. 449.

Barrie C., Ho J. (2021) academictwitteR: An R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software*, vol. 6, no 62, p. 3272.

Bayad A., Sakin D., Cesur S. (2020) Orta Öğretim Öğrencilerinde Barışın Sosyal Temsilleri. *Türk Psikoloji Dergisi*, vol. 36, no 88.

Benoit K. (2020) Text as data: An overview. *The SAGE Handbook of Research Methods in Political Science and International Relations* (L. Curini & R. Franzese eds.), London: Sage, pp. 461–497.

Benoit K., Watanabe K., Wang H., Nulty P., Obeng A., Müller S., Matsuo A. (2018) quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, vol. 3, no 30, p. 774.

Bharathi Mohan G., Prasanna Kumar R., Parathasarathy S., Aravind S., Hanish K. B., Pavithria G. (2023) Text Summarization for Big Data Analytics: A Comprehensive Review of GPT 2 and BERT Approaches. *Data Analytics for Internet of Things Infrastructure. Internet of Things* (R. Sharma, G. Jeon, Y. Zhang eds.), Cham: Springer.

Boyles J. L., Weber M. S., Borges-Rey E. (2024) In Code We Trust? Assessing Code's Role as a Mediator of Power and Ethics Within Journalistic Practice. *Digital Journalism*, vol. 12, no 7, pp. 914–925.

Cai Q., Lin Y., Yu Z. (2023) Factors Influencing Learner Attitudes Towards ChatGPT-Assisted Language Learning in Higher Education. *International Journal of Human–Computer Interaction*, vol. 40, no 22, pp. 7112–7126.

Cai W. (2022) Environment, Aesthetics, Technology, and Gentrification from Global North to Global South: An Integrated Review of Literature. *Journal of Smart Cities*, vol. 7, no 1, pp. 5-12.

Çakıcı Z. (2020) *Uluslararası dijital medyada bir sosyal temsil olarak Aylan Kurdi haberleri (Yayımlanmamış yüksek lisans tezi),* Galatasaray Üniversitesi.

Calma J. (2023) *Twitter just closed the book on academic research — The Verge*. https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research (accessed 10 July 2024)

Carpenter D., Min W., Lee S., Ozogul G., Zheng X., Lester J. (2024) *Assessing Student Explanations with Large Language Models Using Fine-Tuning and Few-Shot Learning,* pp. 403–413.

Cirhinlioğlu F. G., Aktaş V., Öner Özkan B. (2006) *Sosyal temsil kuramına genel bir bakış*.

Creswell A., White T., Dumoulin V., Arulkumaran K., Sengupta B., Bharath A. A. (2018) Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, vol. 35, no 1, pp. 53–65.

Dahal B., Kumar S. A. P., Li Z. (2019) Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, vol. 9, no 1, p. 24.

Demirel S., Bulur N., Çakıcı Z. (2024a) Utilizing Artificial Intelligence for Text Classification in Communication Sciences: Reliability of ChatGPT Models in Turkish Texts. *Advances in Computational Intelligence and Robotics* (D. Darwish ed.), IGI Global, pp. 218–235.

Demirel S., Kahraman E., Gündüz U. (2024b) A text mining analysis of the change in status of the Hagia Sophia on Twitter: The political discourse and its reflections on the public opinion. *Atlantic Journal of Communication*, vol. 32, no 1, pp. 63–90.

DiMaggio P., Nag M., Blei D. (2013) Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U. S. government arts funding. *Poetics*, vol. 41, no 6, pp. 570–606.

Doise W. (2002) Les représentations sociales: Leçons du passé et défis d'aujourd'hui. *Social Science Information*, vol. 41, no 1, pp. 101–110.

Durkheim E. (2014) *Sosyolojik yöntemin kuralları* (Ö. Doğan, Trans.). Doğu Batı Yayınları.

Duveen G., Lloyd B. (1986) The significance of social identities. *British Journal of Social Psychology*, vol. 25, no 3, pp. 219–230.

Fabian T. (2023) The Cool Runnings Effect: Flexible Citizenship, the Global South, and Transcultural Republics at the Winter Olympic Games. *The Olympic Winter Games at 100*). Routledge, pp. 58-83.

Farr R. M., Markova I. (1995) Professional and lay representations of health, illness and handicap: A theoretical overview. *Representations of Health, Illness and Handicap*, pp. 93–110.

Feuerriegel S., Hartmann J., Janiesch C., Zschech P. (2024) Generative AI. *Business & Information Systems Engineering*, vol. 66, no 1, pp. 111–126.

Flacher D., Graña J. M., Rikap C. (2024) Industry 4.0: a global North, global South perspective. *Revue d'économie industrielle*, no 185, pp. 11-15.

Floridi L., Chiriatti M. (2020) GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, no 30, pp. 681–694.

Fütterer T., Fischer C., Alekseeva A., Chen X., Tate T., Warschauer M., Gerjets P. (2023) ChatGPT in education: Global reactions to AI innovations. *Scientific Reports*, vol. 13, no 1, p. 15310.

Gilson A., Safranek C. W., Huang T., Socrates V., Chi L., Taylor R. A., Chartash D. (2023) How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Medical Education*, no 9, p. e45312.

Golan O., Milliken A. (2024) *Turing's Enigma: Laying the Groundwork for AI*.

Gondwe G. (2023) CHATGPT and the Global South: how are journalists in sub-Saharan Africa engaging with generative AI?. *Online Media and Global Communication*, vol. 2, no 2, pp. 228-249.

Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. (2020) Generative adversarial networks. *Communications of the ACM*, vol. 63, no 11, pp. 139–144.

Gordon R. E. (2022) *Development disrupted: the Global South in the twenty-first century*, Cambridge University Press.

Gräbner-Radkowitsch C., Strunk B. (2023) Degrowth and the Global South: The twin problem of global dependencies. *Ecological Economics*, no 213, pp. 1-10.

Grimmer J., Stewart B. M. (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, vol. 21, no 3, pp. 267–297.

Grootendorst M. (2022) BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv Preprint arXiv:*2203.05794.

Gültekin S. (2022) Social Representation Theory and Tourism. *Routledge Handbook of Social Psychology of Tourism*, Routledge, pp. 68–75

Gündüz U., Demirel S. (2023) Metaverse-related perceptions and sentiments on Twitter: Evidence from text mining and network analysis. *Electronic Commerce Research*. https://doi.org/10.1007/s10660-023-09745-x

Günther E., Quandt T. (2016) Word Counts and Topic Models: Automated text analysis methods for digital journalism research. *Digital Journalism*, vol. 4, no 1, pp. 75–88.

Haelewaters D., Hofmann T. A., Romero-Olivares A. L. (2021) Ten simple rules for Global North researchers to stop perpetuating helicopter research in the Global South. *PLoS Computational Biology*, vol. 17, no 8, pp. 1-10.

Haralambous Y., Lenca P. (2023) Beyond the Semantic Web: Towards an Implicit Pragmatic Web and a Web of Social Representations. *Future Internet*, vol. 15, no 7, p. 239.

Haswell F., Edelenbosch O. Y., Piscicelli L., van Vuuren D. P. (2024) The geography of circularity missions: A cross-country comparison of circular economy policy approaches in the Global North and Global South. *Environmental Innovation and Societal Transitions*, vol. 52, pp. 1-10.

Hutto C., Gilbert E. (2014) VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no 1, pp. 216–225.

Hvitfeldt E., Silge J. (2022) *Supervised machine learning for text analysis in R,* Chapman and Hall/CRC.

Ishida T., Ihsan A. F., Rudawan R. A. (2024) Advancing Global South University Education with Large Language Models. *arXiv preprint arXiv*:2410.07139.

Jang S. H., Jung G. (2021) How does COVID-19 differ from the flu/cold? A study of multilevel information seeking among Korean immigrant women in the US. *American Journal of Health Behavior*, vol. 45, no 4, pp. 665–676.

Jo T. (2019) *Text Mining — Concepts, Implementation, and Big Data Challenge (1st ed.),* Cham: Springer.

Jónsson E., Stolee J. (2015) *An evaluation of topic modelling techniques for twitter,* University of Toronto.

Justo A. M., da Silva Bousfield A. B., Giacomozzi A. I., Camargo B. V. (2020) Communication, social representations and prevention-information polarization on COVID-19 in Brazil. *Papers on Social Representations*, vol. 29, no 2, pp. 4–1.

Kahraman E., Demirel S., Gündüz U. (2023) COVID-19 vaccines in twitter ecosystem: Analyzing perceptions and attitudes by sentiment and text analysis method. *Journal of Public Health*. https://doi.org/10.1007/s10389-023-02078-x

Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F., Gasser U., Groh G., Günnemann S., Hüllermeier E., Krusche S., Kutyniok G., Michaeli T., Nerdel C., Pfeffer J., Poquet O., Sailer M., Schmidt A., Seidel T., … Kasneci G. (2023) ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, no 103, p. 102274.

Keiper M. C. (2023) ChatGPT in practice: Increasing event planning efficiency through artificial intelligence. *Journal of Hospitality, Leisure, Sport & Tourism Education*, no 33, p. 100454.

Klinger U., Kreiss D., Mutsvairo B. (2023) Platforms, Power, and Politics: A Model for an Ever-changing Field. *Political Communication Report*, no 27.

Koonchanok R., Pan Y., Jang H. (2023) *Public Attitudes Toward ChatGPT on Twitter: Sentiments, Topics, and Occupations*. https://doi.org/10.48550/ARXIV.2306.12951

Koonchanok R., Pan Y., Jang H. (2024) Public attitudes toward chatgpt on twitter: Sentiments, topics, and occupations. *Social Network Analysis and Mining*, vol. 14, no 1, p. 106.

Koskinen K., Bonina C., Eaton B. (2019) Digital platforms in the global south: foundations and research agenda. In Information and Communication Technologies for Development. Strengthening Southern-Driven Cooperation as a Catalyst for ICT4D: *15th IFIP WG 9.4 International Conference on Social Implications of Computers in Developing Countries, ICT4D 2019, Dar es Salaam, Tanzania, May 1–3, 2019, Proceedings, Part I 15,* Springer International Publishing, pp. 319-330.

Kumar A., Kumar L. (2024) Navigating the Future: The Ethical, Societal and Technological Implications of Artificial Intelligence. *Journal Homepage: Https://Gjrpublication. Com/Gjrecs*, no 4(02).

Kumar S. (2024) Trends of exports of high technology products from Global South. *Development Cooperation Review*, vol. 7, no 1, p. 51.

Leal D. F., Harder N. L. (2021) Global dynamics of international migration systems across South–South, North–North, and North–South flows, 1990–2015. *Applied Network Science*, no 6, pp. 1-27.

Li W., Haunert J., Knechtel J., Zhu J., Zhu Q., Dehbi Y. (2023) Social media insights on public perception and sentiment during and after disasters: The European floods in 2021 as a case study. *Transactions in GIS*, vol. 27, no 6, pp. 1766-1793.

Li W., Su Z., Li R., Zhang K., Wang Y. (2020) Blockchain-based data security for artificial intelligence applications in 6G networks. *IEEE Network*, vol. 34, no 6, pp. 31–37.

Lin T. C. (2019) Artificial intelligence, finance, and the law. *Fordham L. Rev.*, no 88, p. 531.

Liu B. (2020) *Sentiment analysis: Mining opinions, sentiments, and emotions* (Second edition), Cambridge university press.

Lucas C., Nielsen R. A., Roberts M. E., Stewart B. M., Storer A., Tingley D. (2015) Computer-Assisted Text Analysis for Comparative Politics. *Political Analysis*, vol. 23, no 2, pp. 254–277.

Malczewski E. (2019) Durkheim and the Nation. *İstanbul University Journal of Sociology*, vol. 39, no 1, pp. 41–64.

Martikainen J. (2020) *Visual representations of teachership: A social representations approach*.

McCarthy J., Minsky M. L., Rochester N., Shannon C. E. (2006) A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI Magazine*, vol. 27, no 4, pp. 12–12.

Meyer J. (2018) Restructuring the Christian Fatherhood Model: A practical theological investigation into the 'male problematic'of father absence. *HTS: Theological Studies*, vol. 74, no 1, pp. 1–11.

Molosi-France K., Makoni S. (2020) A partnership of un-equals: global south–north research collaborations in higher education institutions. *Modern Africa: Politics, History and Society*, vol. 8, no 2, pp. 9-24.

Moreira L., Paiva J. C., Morais C. (2021) The social representations of the Internet: A systematic review of literature towards a groundbreaking research agenda. *Papers on Social Representations*, vol. 30, no 1, p. 1–1.

Moscovici S. (1963) Attitudes and Opinions. *Annual Review of Psychology*, vol. 14, no 1, pp. 231–260.

Mudaly R., Chirikure T. (2023, October) STEM education in the Global North and Global South: Competition, conformity, and convenient collaborations. *Frontiers in Education*, vol. 8, p. 1144399. Frontiers Media SA.

Müller M. (2018) In Search of the Global East: Thinking between North and South. *Geopolitics*, vol. 25, no 3, pp. 734–755.

Nedbal C., Naik N., Castellani D., Gahuar V., Geraghty R., Somani B. K. (2023) ChatGPT in urology practice: Revolutionizing efficiency and patient care with generative artificial intelligence. *Current Opinion in Urology*. https://doi.org/10.1097/MOU.0000000000001151

Nisch S. (2024) Public opinion about Finland joining NATO: Analysing Twitter posts by performing natural language processing. *Journal of Contemporary European Studies*, vol. 32, no 1, pp. 272-290.

Novikova I., Berezina E., Sachkova M., Dvoryanchikov N., Novikov A., Bovina I. (2024) To Be Scared or Not to Be Scared: Social Representations of COVID-19 in Young People (A Cross-Cultural Study). *Social Sciences*, vol. 13, no 1, p. 62.

Ogoro M., Minescu A., Moriarty M. (2022) Cultural Identity in Bicultural Young Adults in Ireland: A Social Representation Theory Approach. *Social Sciences*, vol. 11, no 6, p. 230.

Oliński M., Krukowski K., Sieciński K. (2024) Bibliometric Overview of ChatGPT: New Perspectives in Social Sciences. *Publications*, vol. 12, no 1, p. 9.

Öner B. (2002) SOSYAL TEMSİLLER. *Kriz Dergisi*, 029–035. https://doi.org/10.1501/Kriz_0000000179

Pang B., Lee L. (2008) Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, vol. 2, no 1–2, pp. 1–135.

Paul K., Milmo D. (2022, October 28) Elon Musk completes Twitter takeover and 'fires top executives.' *The Guardian*. https://www.theguardian.com/technology/2022/oct/27/elon-musk-completes-twitter-takeover (accessed 10 July 2024)

Pineda P., Mishra S. (2023) The semantics of diversity in higher education: differences between the Global North and Global South. *Higher Education*, vol. 85, no 4, pp. 865-886.

Polli G. M., Camargo B. V. (2015) Social Representations of the Environment in Press Media. *Paidéia (Ribeirão Preto)*, vol. 25, no 61, pp. 261–269.

Portes Virginio F., Stewart P., Garvey B. (2023) Unpacking super-exploitation in the 21st century: The struggles of Haitian workers in Brazil. *Work, Employment and Society*, vol. 37, no 4, pp. 897-915.

Pozzi F. A., Fersini E., Messina E., Liu, B. (2017) Challenges of Sentiment Analysis in Social Networks. *Sentiment Analysis in Social Networks*. Elsevier, pp. 1–11.

Qi W., Pan J., Lyu H., Luo J. (2024) Excitements and concerns in the post-ChatGPT era: Deciphering public perception of AI through social media analysis. *Telematics and Informatics*, no 92, p. 102158.

Qin X., Huang M., Ding J. (2024) AITurk: Using ChatGPT for social science research. SSRN. https://doi.org/10.2139/ssrn.4922861

Quinn K. M., Monroe B. L., Colaresi M., Crespin M. H., Radev D. R. (2010) How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science*, vol. 54, no 1, pp. 209–228.

R Core Team (2023) R: A Language and Environment for Statistical Computing [Computer software]. *R Foundation for Statistical Computing*. https://www.R-project.org/

Rabitz F., Telešienė A., Zolubienė E. (2021) Topic modelling the news media representation of climate change. *Environmental Sociology*, vol. 7, no 3, pp. 214–224.

Ramage D., Rosen E., Chuang J., Manning C. D., McFarland D. A. (2009) *Topic modeling for the social sciences,* vol. 5, no 27, pp. 1–4.

Ribeiro F. N., Araújo M., Gonçalves P., André Gonçalves M., Benevenuto F. (2016) SentiBench — A benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, vol. 5, no 1, p. 23.

Rovamo H., Sakki I. (2024) Lay representations of populism: Discursive negotiation of naturalized social representation. *Journal of Community & Applied Social Psychology*, vol. 34, no 1, p. e2755.

Ruan T., Lv Q. (2022) Public perception of electric vehicles on reddit over the past decade. *Communications in Transportation Research*, no 2, p. 100070.

Saleh S. N., McDonald S. A., Basit M. A., Kumar,S., Arasaratnam R. J., Perl T. M., Lehmann C. U., Medford R. J. (2023) Public perception of COVID-19 vaccines through analysis of Twitter content and users. *Vaccine*, vol. 41, no 33, pp. 4844-4853.

Schembri P., Petit O. (2009) Clean technology transfers and North-South technological gap: An important issue for environmental policies. Économie *internationale*, vol. 120, no 4, pp. 109-129.

Silge J., Robinson D. (2017) *Text mining with R: A tidy approach,* O'Reilly Media, Inc.

Singh H., Tayarani-Najaran M.-H., Yaqoob M. (2023b) Exploring Computer Science Students' Perception of ChatGPT in Higher Education: A Descriptive and Correlation Study. *Education Sciences*, vol. 13, no 9, p. 924.

Singh S. K., Kumar S., Mehra P. S. (2023a) *Chat gpt & google bard ai: A review,* pp. 1–6.

Smejkalová T. (2024) Legal Concepts as Social Representations. *International Journal for the Semiotics of Law-Revue Internationale de Sémiotique Juridique*, vol. 37, no 1, pp. 165–188.

Souza L. G. S., O'Dwyer E., Coutinho S. M. dos S., Chaudhuri S., Rocha L. L., Souza L. P. de. (2021) Social representations and ideology: Theories of common sense about COVID-19 among middle-class Brazilians and their ideological implications. *Journal of Social and Political Psychology*, vol. 9, no 1, pp. 105–122.

Sovacool B. K. (2023) Expanding carbon removal to the Global South: Thematic concerns on systems, justice, and climate governance. *Energy and Climate Change*, no 4, pp. 1-10.

Sultana F. (2023) Whose growth in whose planetary boundaries? Decolonising planetary justice in the Anthropocene. *Geo: Geography and Environment*, vol. 10, no 2, p. e00128.

Tarisayi K. S. (2024) ChatGPT use in universities in South Africa through a socio-technical lens. *Cogent Education*, vol. 11, no 1, pp. 2295654.

Tian S., Jin Q., Yeganova L., Lai P.-T., Zhu Q., Chen X., Yang Y., Chen Q., Kim W., Comeau D. C., Islamaj R., Kapoor A., Gao X., Lu Z. (2024) Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Briefings in Bioinformatics*, vol. 25, no 1, Article bbad493.

Tong J., Zuo L. (2021) Social Representations of the EU Referendum on Twitter. *The Brexit Referendum on Twitter: A Mixed-Method, Computational Analysis*. Emerald Publishing Limited, pp. 21–32.

Tureček P., Kleisner K. (2022) Symptomic Mimicry Between SARS-CoV-2 and the Common Cold Complex. *Biosemiotics*, vol. 15, no 1, pp. 61–66.

Turing A. M. (1950) I. — COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, *LIX, no* 236, pp. 433–460.

Verma M. (2023) Artificial intelligence role in modern science: Aims, merits, risks and its applications. *Artificial Intelligence*, vol. 7, no 5.

Wang K., Variengien A., Conmy A., Shlegeris B., Steinhardt J. (2022) Interpretability in the wild: A circuit for indirect object identification in GPT-2 small. *arXiv Preprint*, arXiv:2211.00593. https://doi.org/10.48550/arXiv.2211.00593

Wang W., Siau K. (2019) Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management (JDM)*, vol. 30, no 1, pp. 61–79.

Welbers K., Van Atteveldt W., Benoit K. (2017) Text Analysis in R. *Communication Methods and Measures*, vol. 11, no 4, pp. 245–265.

Wiedemann G. (2016) *Text mining for qualitative data analysis in the social sciences,* Springer Berlin Heidelberg.

Woschank M., Rauch E., Zsifkovits H. (2020) A review of further directions for artificial intelligence, machine learning, and deep learning in smart logistics. *Sustainability*, vol. 12, no 9, p. 3760.

X/Twitter: Global audience 2024. (2024) Statista. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ (accessed 10 July 2024)

Yin H., Song X., Yang S., Li J. (2022). Sentiment analysis and topic modeling for COVID-19 vaccine discussions. *World Wide Web*, vol. 25, no 3, pp. 1067–1083.

Yu H. (2023) Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology*, vol. 14, p. 1181712.

Zengele T., Pitsoe V. J. (2023) Moscovici and Schon on Quality Assurance in Open Distance Learning. *Available at SSRN 4497228*.

Zhai W., Hu W., Yuan Z., Li Y. (2024) Examining disaster resilience perception of social media users during the billion-dollar hurricanes. *Natural Hazards*, vol. 120, no 1, pp. 701-727.

Zhang L., Liu B. (2017) Sentiment Analysis and Opinion Mining. *Encyclopedia of Machine Learning and Data Mining* (C. Sammut, G. I. Webb eds.), US Springer, pp. 1152–1161.

Zhao J., Liu K., Xu L. (2016) Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. *Computational Linguistics*, vol. 42, no 3, pp. 595–598.

Zheng X., Zhang C., Woodland P. C. (2021, December) Adapting GPT, GPT-2 and BERT language models for speech recognition. *2021 IEEE Automatic Speech Recognition and Understanding Workshop* (ASRU). IEEE. pp. 162–168.

Zhong B. (2023). Going beyond fact-checking to fight health misinformation: A multi-level analysis of the Twitter response to health news stories. *International Journal of Information Management*, no 70, p. 102626.

Zolnoori M., Huang M., Patten C. A., Balls-Berry J. E., Goudarzvand S., Brockman T. A., Sagheb E., Yao L. (2021) Mining news media for understanding public health concerns. *Journal of Clinical and Translational Science*, vol. 5, no 1, p. e1.

Zou W., Li J., Yang Y., Tang L. (2023) Exploring the Early Adoption of Open AI among Laypeople and Technical Professionals: An Analysis of Twitter Conversations on #ChatGPT and #GPT3. *International Journal of Human–Computer Interaction*, p. 1–12.

Zubiaga A., Procter R., Maple C. (2018) A longitudinal analysis of the public perception of the opportunities and challenges of the Internet of Things. *PLOS ONE*, vol. 13, no 12, p. e0209472.

Социальное восприятие искусственного интеллекта в Twitter: сравнительное исследование стран Глобального Юга и Глобального Севера

Садеттин Демирель
Доцент Факультета новых медиа и коммуникаций Университета Ускюдар
Адрес: Алтунизаде, ул. Кысыклы, 16-2, 34662, Ускюдар
Стамбул, Турция
E-mail: sadettindemirel1@gmail.com

Зиндан Чакыджи
Доцент Факультета новых медиа и коммуникаций Университета Ускюдар
Адрес: Алтунизаде, ул. Кысыклы, 16-2, 34662, Ускюдар
Стамбул, Турция
E-mail: sadettindemirel1@gmail.com

Неслихан Булур
Научная сотрудница Факультета новых медиа и коммуникаций Университета Ускюдар
Адрес: Алтунизаде, ул. Кысыклы, 16-2, 34662, Ускюдар
Стамбул, Турция
E-mail: neslihan.bulur@uskudar.edu.tr

Это исследование изучает социальное восприятие ChatGPT в Twitter, сравнивая подходы в странах Глобального Юга (Индия и Турция) и Глобального Севера (Великобритания и США). Используется подход «текст как данные», включающий анализ тональности высказываний, анализ текста и тематическое моделирование твитов на английском языке, содержащих слово «ChatGPT», собранных в период с 1 января по 1 июня 2023 года. Для анализа тональности применяется словарь тональности VADER, а для текстового анализа и тематического моделирования используются n-граммы и BERTopic соответственно. Анализ тональности выявляет в целом позитивное восприятие ChatGPT в обоих регионах, хотя также выражаются опасения по поводу возможных негативных последствий. Текстовый

анализ и тематическое моделирование показывают, что дискуссии сосредоточены вокруг крупных технологических компаний и ведущих фигур, при этом страны Глобального Севера больше акцентируют внимание на отраслевом применении и технических аспектах, тогда как Глобальный Юг оценивает ChatGPT в контексте местных языков и культур. Результаты исследования демонстрируют, что социокультурные различия и межрегиональные уровни технологического развития отражаются и в социальном восприятии ChatGPT. Исследование вносит вклад в понимание раннего восприятия генеративного искусственного интеллекта и подчеркивает важность учета вопросов конфиденциальности, разработки культурно совместимых решений и изучения потенциального влияния ChatGPT в различных областях.
Ключевые слова: социальное восприятие, искусственный интеллект, ChatGPT, глобальный юг, глобальный север, Twitter/X

doi: 10.17323/1728-192x-2024-4-80-112

Aadhaar, AI, and Identity: Negotiating Power and Surveillance in the Global South

Nilanjana Sinha
Associate Professor, NSHM Business School
Address: 124 B L Saha Road, Kolkata — 700 053, West Bengal, India
E-mail: nilanjana.sinha@gmail.com

Originally intended as a tool to streamline governance and facilitate inclusion, Aadhaar has developed into an amalgam that engenders deeply concerning ramifications for Indian society. Enumerated as an inquiry into a technology-based public administration system, Aadhaar unveils a convoluted mechanism that poses pertinent questions regarding surveillance, governance, and identity. The study delves into the paradox of Aadhaar: how it acts as both a tool of governance and an instrument of identification surveillance, utilizing a combination of interviews, processing data analysis, and post-colonialism to unpack its workings within Indian society.

Even while Aadhaar promises to make benefits accessible to people easily, practical possibilities present a challenge due to the presence of duplicate ID cards or living-active accounts of persons who died, for instance. This development consequently raises fears about mass surveillance and the privatization of public functions furthered by private companies such as OnGrid and Khosla Labs, which calls into great question issues of accountability and ethical consideration. In comparison with other global regimes such as the GDPR or the Social Security System, Aadhaar lacks strict mechanisms to protect an individual's privacy and ensure accountability and all necessary levels of transparency.

Aadhaar's surveillance mechanisms, based on Foucault's concept of the panopticon, normalize the state's power to monitor and control its citizens through biometric tracking. This transcends inclusion into the realm of a new form of identity governance where those with lesser chances for acceptance-think migrants, poor people, and the unbanked-are again in a slightly more analytically navigable environment of exclusion. Hybridity, in the work of Bhabha, enables another way of comprehending the workings of Aadhaar in the post-colonial context: it is fluid and must be constantly redefined by surveillance. The critical post-colonial analysis outlined by Edward Said delineates how the Aadhaar has gone to mimic colonial systems of categorization while further fortifying power hierarchies and systematic, institutional exclusion.

As Aadhaar increasingly shapes Indian identity, it does so not merely as a tool of inclusion, but as a system that systematically places citizens under surveillance — making identity synonymous with scrutiny and regulation, normalizing the power to monitor and control without consent or choice.

Keywords: Aadhaar, Biometric Identification, Surveillance, Identity, Post-Colonial Theory.

Introduction

The rise of highly concentrated AI in the Global North has driven innovation but exacerbated disparities in the Global South, widening the North-South divide. Kulesz (Kulesz, 2022: 56) warns of a "dual divide, technological and creative", threatening equitable growth. While AI advancements benefit agriculture, healthcare, and education (Okolo, 2023), challenges like data governance, algorithmic bias, and privatized local research hinder progress (Sharma, Ramann, 2023). Governance models from the Global North often replicate neo-colonial practices, perpetuating structural inequalities (Basu, Omo-

tuboraa, 2024). AI-powered surveillance undermines civil liberties, fostering cultural hegemony and alienating local contexts, creating a sense of "asphyxia" (Kulesz, 2017).

The Global North's dominance in data control exacerbates these imbalances, creating economic dependency in the Global South. India exemplifies this "data colonialism" through its Aadhaar system, a biometric identification initiative showcasing both the promises and risks of AI-driven systems. Aadhaar, managed by the Unique Identification Authority of India (UIDAI) under the Aadhaar Act of 2016, provides Indian residents with a 12-digit unique identity number (UID) based on biometric and demographic data. Although Aadhaar is not an AI system, it is a foundational infrastructure for AI-powered applications like biometric authentication, fraud detection, and predictive analytics. By March 2023, Aadhaar had issued 136.65 crore IDs, supported by 64,003 enrollment and update centers. Aadhaar updates rose from 0.55 lakh in 2012-13 to 2001.5 lakh in 2022-23, while authentications surged from 0.24 crore to 2291.97 crore during the same period. Similarly, e-KYC transactions grew from 89.1 crore in 2019-20 to 322.97 crore in 2022-23 (UIDAI Annual Report, 2022-23). Moreover, 77.03 crore Aadhaar numbers were linked to bank accounts by March 2023, up from 6.70 crore in 2014, underlining Aadhaar's role in financial inclusion (UIDAI Annual Report, 2022-23).

Aadhaar, once envisioned as a voluntary ID system, has transformed into a near-essential passport to life in India. Although the Aadhaar Act (2016) initially limited its mandatory use to subsidies like LPG and welfare programs under Section 7 (UIDAI Guidelines, 2016), its influence has silently crept into every corner of daily life. Aadhaar often becomes an unspoken but unavoidable requirement, skirting its legal boundaries (Singal, 2022). The mandatory Aadhaar-PAN linkage for tax filings under Section 139AA has anchored it firmly within the financial system. Meanwhile, welfare schemes like MGNREGA and school midday meals now hinge on Aadhaar verification. Yet, countless marginalized individuals face exclusion due to biometric failures and system errors (Press Trust of India, 2019). Critics warn that Aadhaar, once a symbol of inclusion and opportunity, risks evolving into a tool of control — deepening inequalities and fortifying state power. What began as a promise of empowerment now walks a fine line, raising questions about access, equity, and agency. As The Wire noted in May 2024:

"Over the last 10 years, the number of things we must possess an Aadhaar number to access in India has only gone up, including (but not limited to): midday meals; filing income tax returns; University Grants Commission fellowships; compensation schemes; MGNREGA work; a number of government schemes; death registration; and even deep-sea fishing. Despite carefully maintaining that it is a 'voluntary' program, the regime has slowly but surely ensured that Aadhaar is mandatory in all the ways that matter."

Research Objective

This study examines how Aadhaar, India's AI-driven identity system, influences identity and governance, focusing on privacy, equity, and human rights in a surveillance framework. It explores three key questions:

- How does Aadhaar shape individual and collective identities in post-colonial India?
- How does Aadhaar reinforce or challenge colonial governance, surveillance, and classification?
- How does Aadhaar's integration into life shape identity through disciplinary power and normalization?

Using post-colonial and post-structuralist theories, particularly Homi Bhabha's hybridity, Edward Said's cultural critique, and Foucault's ideas on power, the study combines theoretical insights and empirical analysis to assess Aadhaar's impact on governance, identity, and ethics.

The research combines a literature review of key theorists with empirical analysis, focusing on Aadhaar's infrastructure, its influence on identity formation, and ethical implications. This approach provides a nuanced understanding of how Aadhaar reshapes governance and societal behavior.

Theoretical Context: Said, Bhabha, and Foucault on Power and Identity

Drawing on postcolonial and poststructuralist theories, the research critically examines whether such systems perpetuate or disrupt colonial power dynamics and their impact on the diversity of global cultural identities.

Post-Colonial Theory

Orientalism and the Power Dynamics of Representation

Edward Said's "Orientalism" offers a sharp critique of how the West has historically constructed and perceived the East; a process rooted in the need to justify colonial ambitions. Said argues that Western representations (the Occident) often depict Eastern civilizations (the Orient) as the Other. This binary division marginalizes the Orient and upholds Western dominance, despite its artificial and contrived nature. The Western literary and cultural canon has long supported and perpetuated misleading images that reinforce Western dominance. Said's analysis, which echoes Karl Marx's assertion that "They cannot represent themselves; they must be represented," aligns with Gayatri Chakravorty Spivak's exploration in "Can the Subaltern Speak?" — both highlighting how Western narratives and dominant discourses disempower and silence marginalized voices. This concept ties into the broader idea that knowledge and power are inextricably linked (Loomba, 1998). In essence, Said's "Orientalism" challenges the enduring colonial narrative by exposing the complex power dynamics and systematic misrepresentations embedded in Western depictions of the East.

Postcolonial theories, as articulated by Boehmer (1995), offer a new interpretation of the traditional dynamic between the self and the Other. Orientalism creates its own Other to strengthen its identity and superiority, situating the Orient as a concealed self.

This biased representation, seen in films like Danny Boyle's "Slumdog Millionaire," has historically justified Western intervention and dominance. Said maintains that Orientalism inaccurately portrays Eastern societies, perpetuating ideas of Western superiority and colonial agendas (Said, 2001).

Bhabha's Challenge to Traditional Colonial Discourses

Homi Bhabha's concept of hybridity, linked to mimicry, offers a nuanced perspective on colonial discourse. Mimicry involves colonized individuals adopting traits of their colonizers, resulting in hybrid identities that blur distinctions between colonizer and colonized. This disrupts traditional binaries and creates spaces for negotiation and resistance within colonial power structures. Bhabha argues that colonialism influences the identities of both the colonized and the colonizers, challenging static notions of self and other. His concept of "metonymy of presence" shows how mimicry subverts colonial authority through ongoing identity negotiation (McLeod, 2000: 32-33).

Frantz Fanon argues that colonial powers impose negative identities on the colonized, forcing them to see themselves as objects rather than subjects, effectively denying them the right to self-definition (McLeod, 2000: 20). Similarly, Homi Bhabha, in The Location of Culture (1994), discusses how colonial discourse uses stereotypes to portray the colonized as degenerate, justifying their control. He notes that stereotyping involves a complex mix of projection and introjection, creating a veneer of "official" knowledge while concealing underlying anxieties" (Bhabha, 1986: 169). Bhabha also observes that "to be Anglicized is emphatically not to be English," highlighting the inherent ambivalence in colonial mimicry (Bhabha, 1994: 87).

Understanding Power in Modern Societies

Michel Foucault's theory of power, as discussed by Gaventa (2003) and in Foucault's own writings, redefines power in modern societies as diffuse, embodied, and discursive, asserting that "power is everywhere" and "comes from everywhere" (Foucault, 1998: 63). Power, in Foucault's view, is not a static force or structure but functions as a 'regime of truth,' constantly shifting and evolving through accepted knowledge and discourse. He introduces the concept of governmentality, or the "conduct of conduct," which refers to the strategies and methods used to guide and control behavior, emphasizing the difference between the practical aspects of governance and the underlying ideas that make these practices effective (Li, 2007).

Foucault critiques traditional power structures by introducing the concept of "disciplinary power," which contrasts with the coercive power of earlier feudal states. In modern times, this power is evident in systems emphasizing surveillance and normalization rather than overt force (LotusArise, 2024). Foucault (1998) notes the efficiency and subtlety of these new techniques, stating, "These new techniques were not only much more efficient and much less extravagant... but above all this period saw the creation of a new

'economy' of power" (Foucault, 1998: 100-101). He challenges the view of power as merely repressive, noting, "If power has only ever been repressive if all it does is to say no, do you think that one would obey it? What makes power effective... is not simply that it is felt as a power that says no, but that in fact, it produces things, it produces pleasure, it creates knowledge" (Foucault, 1998: 63). This perspective positions power as a productive network that transcends simple punitive functions.

Foucault's theory shares common ground with Said's and Bhabha's analyses in addressing how power constructs and perpetuates knowledge and identity, whether through colonial representations (Said), hybrid identities (Bhabha), or disciplinary mechanisms (Foucault). However, while Said focuses on the binary opposition and misrepresentations in colonial discourse, Bhabha emphasizes the fluidity and subversive potential within colonial encounters, and Foucault shifts the focus to the mechanisms of power in modern administrative and social systems. Foucault's notion of power as diffuse and productive contrasts with Said's more static critique of colonial representations and Bhabha's emphasis on hybrid identities emerging from cultural interactions. Foucault's approach is concerned with how power operates through everyday practices and discourses to produce and regulate behavior, rather than through explicit representations or identity negotiations (Meloni, 2023).

Methodology

This study employs decolonizing research methods, inspired by Jacqueline M. Quinless's Decolonizing Data: Unsettling Conversations about Social Research Methods (2022). Quinless advocates for anti-racist, anti-oppressive approaches that prioritize marginalized voices and integrate diverse epistemologies and ethical practices. This methodology seeks to amplify historically silenced perspectives and de-center traditional, top-down research frameworks, focusing on the lived experiences of those impacted by Aadhaar, India's biometric identification system.

Study Subject. Aadhaar, integrating biometrics with unique IDs, governs citizen-state interactions and shapes social inclusion and exclusion. This study critically explores its goals, implementation, and impact on marginalized communities, focusing on how surveillances shape identity formation.

Research Design. The study employs a mixed-methods design (MMR), integrating quantitative and qualitative data. Following Greene et al. (2004), this approach ensures triangulation, complementarity, and expansion for comprehensive analysis. Incorporating participatory research, it actively involves marginalized groups impacted by Aadhaar, aligning with Quinless's decolonizing approach to foster inclusive and ethical inquiry.

Data Collection. The data collection strategy combines secondary and primary sources to build a robust and nuanced understanding of Aadhaar's social and political impact.

Secondary Data. A systematic review of secondary sources was conducted, drawing from a range of academic papers, policy documents, government reports, legal frameworks, and media coverage. Key references included documents from the Unique Iden-

tification Authority of India (UIDAI), reports from think tanks, and articles from major newspapers such as The Hindu and The Indian Express. These sources provided insights into Aadhaar's objectives, implementation strategies, regulatory frameworks, and the broader power dynamics involved. Databases like Google Scholar and government archives were also consulted to access diverse and credible perspectives on Aadhaar's impact. The purpose of this secondary data was to offer contextual understanding, analyze theoretical perspectives, and explore Aadhaar's socio-political implications. The systematic review helped synthesize these sources to frame qualitative data and cross-check findings from primary research.

Primary Data

Semi-Structured Interviews. To understand the multifaceted impact of Aadhaar, a series of interviews were conducted with key stakeholders in Kolkata and its suburban areas, each providing unique perspectives on the system's implementation and its social consequences. The research began with interviews with two policymakers directly involved in Aadhaar's rollout, offering insights into the governance and regulatory challenges shaping the system. Additionally, five academic experts from diverse fields were interviewed, raising concerns about Aadhaar's potential to reinforce inequalities, particularly for marginalized communities.

To capture the lived experiences of those most affected, thirty interviews were conducted with individuals from marginalized groups, including religious minorities, low-income families, uneducated women, manual laborers, senior citizens, migrants, and refugees. These participants discussed the technical errors, bureaucratic hurdles, and access issues they faced in using Aadhaar for essential services.

The interview process was conducted with care to ensure ethical standards, including assuring participants of confidentiality and data usage for academic research. Graduate students from the media and animation department helped in shooting and editing the interviews, ensuring high-quality data collection. The process spanned 12 weeks to accommodate challenges in participant recruitment, consent, and the time required for detailed interviews and editing.

Through these interviews, the study gained a nuanced understanding of Aadhaar's impacts, revealing both its potential benefits and significant challenges for vulnerable populations.

Focus Group Discussions. A focus group comprising academicians and student researchers was convened to analyze emerging themes and critically engage with diverse perspectives on Aadhaar. This collaborative approach helped refine the study's analytical lens.

Data Formats. Semi-structured interviews were the primary format for collecting qualitative data, complemented by personal stories, visuals, and narratives to provide a rich, human-centered view of Aadhaar's effects. These materials were crucial in portraying the lived realities of participants and their interactions with the system.

Quantitative Data. Quantitative data includes statistics, figures, and metrics that offer numerical insights into various aspects of the Aadhaar system, such as enrollment rates, usage statistics, and exclusion metrics. This data is sourced from publicly available reports, government databases, and UIDAI statistical data, providing hard evidence on Aadhaar's usage and reach. The quantitative data was used to contextualize and complement qualitative findings, offering a comprehensive understanding of Aadhaar's impact from both statistical and narrative perspectives.

Analytical Framework

This study uses interdisciplinary theories, including Quinless (2022), Bhabha (1994), Said (1978), and Foucault's (1991) biopolitics, to analyze Aadhaar's role in governance and surveillance within a postcolonial context. These frameworks explore how Aadhaar reinforces or challenges colonial legacies of control and exclusion. The research combines qualitative thematic coding of interview transcripts with secondary quantitative data to examine themes like exclusion, privacy, and social identity. A convergent design integrates these methods, creating a comprehensive narrative that links individual experiences to broader systemic trends.

Ethical Considerations

Ethical guidelines were strictly adhered to throughout the study. Informed consent was obtained from all interviewees, ensuring transparency and voluntary participation. Participant confidentiality and privacy were maintained at every stage of the research. Special care was taken to honor the dignity and perspectives of marginalized groups, ensuring their voices were central to the analysis.

By combining, Quinless's decolonizing methodologies with mixed-methods research, the study reimagines Aadhaar narratives, prioritizing marginalized voices and critiquing the power structures behind surveillance systems. Quinless's call for "unsettling conversations" in social research (2022, p. 119) guides this approach, promoting ethical, inclusive, and equitable practices grounded in respect and integrity.

Empirical Insights on Aadhaar: Identity, Power, and Resistance

Launched in 2009 under Nandan Nilekani's leadership, Aadhaar sought to unify fragmented identity systems with a biometric-based Unique Identification Number (UID), aiming to simplify access to government services. Costing an estimated INR 60,000-70,000 crores, Aadhaar operates under the Unique Identification Authority of India (UIDAI) and the Aadhaar Act, 2016. The Act explicitly states its boundaries: Clause 9 of Chapter III clarifies that "The Aadhaar number or authentication thereof shall not, by itself, confer any right of, or be proof of, citizenship or domicile." Similarly, Clause 2(v) defines Aadhaar as a tool for resident identity, not citizenship. As The Hindu em-

phasized, "Aadhaar cards are a proof of identity, not of citizenship or date of birth" (Deep, 2024).

Initially envisioned to improve the Public Distribution System (PDS), Aadhaar has evolved into a de facto national identity system. Despite privacy debates and initial legal challenges, it became effectively mandatory for accessing services by 2017. Commentators like O'Callahan (2020) have called it "a biometrically secured national identification system" central to India's digital identity efforts. While transformative, Aadhaar's journey has raised critical concerns about exclusion, privacy, and the balance of power.

A former UIDAI program director explained Aadhaar's rationale:

"With approximately 3.7 lakh crores allocated for subsidies, yet only 19 paise of every rupee reaching beneficiaries, Aadhaar was designed to enhance transparency, reduce duplication, and streamline welfare services. It aligns demographic and economic data to reach over 56% of the population living below the poverty line and addresses challenges like the 20% homeless population. Aadhaar has notably reduced leakage in benefit distribution."

Before Aadhaar, studies showed only 27% of funds reached intended beneficiaries due to intermediary corruption (Singh, 2017; Henne, 2019). Today, Aadhaar is integral to programs like the Public Distribution System (PDS), direct benefit transfers (DBTs), and the Pradhan Mantri Jan Dhan Yojana (PMJDY).

A senior service provider highlighted its extensive adoption:

"Aadhaar now covers 98% of India's population. Initiatives like Jan Dhan Yojana have increased bank accounts from 8 crores to 18 crores. Aadhaar-enabled e-KYC is widely adopted by banks, mutual funds, and services like Bharat Matrimony. New applications include e-boarding, home voting, and e-credit cards. It supports attendance tracking in government offices, schools, and midday meal programs, while also helping estimate population metrics in the absence of recent census data. Its success has inspired similar systems in Estonia, Saudi Arabia, and Sri Lanka."

Aadhaar has also become foundational for artificial intelligence (AI) integration in governance, supporting fraud detection, predictive analytics, and efficient service delivery. During COVID-19, its integration with CoWIN facilitated AI-powered vaccine tracking. Aadhaar also streamlines financial transactions and education-related identity verification. According to McKinsey, such advancements could contribute up to 13% to India's GDP by 2023.

However, these benefits come with risks. A former Aadhaar director noted:

"Aadhaar's centralized biometric database processes 14,000 records per second, enabling AI-powered governance. But its structure heightens vulnerabilities, exposing data to identity theft, fraud, and misuse, unlike decentralized systems in Europe. This raises critical concerns about privacy, data security, and ethical AI use." Despite safeguards like Clause 30 of the IT Act, 2000, which designates biometric data as "sensitive," high-profile breaches reveal systemic flaws. Maharashtra reported Aadhaar-PAN data leaks during IT uploads, and MS Dhoni's Aadhaar details were publicly disclosed by CSC e-Governance, resulting in its blacklisting (Times of India, 2017). Breaches in Jharkhand's and Kerala's

pension schemes exposed sensitive data due to programming errors (Indian Express, 2017). UIDAI has barred Axis Bank, eMudhra, and Suvidhaa Infoserve over fraudulent activities, but issues persist (Raju, Singh, and Khatter, 2017). Section 33 (Clause 8) of the Aadhaar Act, 2016, which permits paid access to demographic data, heightens vulnerabilities for marginalized groups (Hennen, 2019). These incidents highlight the need for stronger data protection laws and better governance as Aadhaar integrates with emerging technologies.

The next section will analyze Aadhaar's objectives and limitations, focusing on security flaws, data breaches, impacts on marginalized groups, and surveillance concerns, while comparing its privacy standards to global benchmarks.

(a) Aadhaar's Dual-Edged Sword: Promise vs. Pitfalls

Legal expert Usha Ramanathan has criticized the Aadhaar project for prioritizing enrollment over practical implementation, raising significant concerns about residents' privacy and liberty (Mukherjee, Nayar, 2011; Greenleaf, 2010). While enrollment is technically voluntary, it is essential to access government benefits, sparking fears about the potential misuse of personal data (Kaushik, 2010) and widespread distrust in the government's data management (UIDAI Report, 2010). Issues such as fake identities, fraud, and biometric spoofing further complicate the situation (Brindaalakshmi, 2013; Pati, Kumar, Jain, 2015). While biometric systems address identity fraud, they do not solve issues related to eligibility or quantity fraud (Khera, 2019).

Recent reports have highlighted significant discrepancies in Aadhaar management, such as the number of Aadhaar cards exceeding state populations due to errors, migration, or the failure to deactivate cards of deceased individuals. Despite acknowledging these concerns, the government has not yet implemented a system to deactivate Aadhaar cards of the deceased (Table 1). In addition, surges in Aadhaar registrations, particularly in regions near borders and states like Uttar Pradesh and Bihar, raise alarms that AI systems relying on Aadhaar data may either approve fraudulent transactions or deny rightful benefits. The proliferation of counterfeit Aadhaar cards also presents risks to election integrity (Sarma, 2023). The issue is closely related to AI because AI systems, which rely on Aadhaar data for identity verification, could be compromised by these discrepancies. For instance, AI algorithms may incorrectly approve fraudulent transactions or deny legitimate benefits if they rely on incomplete or outdated Aadhaar data, such as cards that belong to deceased individuals or those issued erroneously. Additionally, AI-powered systems used in elections for voter verification could be tricked by counterfeit Aadhaar cards, undermining election integrity. These risks highlight the need for more robust data management and AI systems that can detect and prevent such fraud.

A report by Lingamgunta Nirmitha Rao (2023) indicates that UIDAI plans to implement a deactivation mechanism for deceased individuals. While updates to Aadhaar can now be done free of charge through the 'myAadhaar' portal (March 15 to June 14, 2023),

Table 1: States that boosted Aadhar cards (between 31st Dec 2021 and 28 Feb 2023)

States and Union Territories	Big Boost (%)
Meghalaya	24.5
Assam	9.3
Mizoram	3.8
Uttar Pradesh	3.6
Manipur	3.4
West Bengal	3.3
Ladakh	3.1
Bihar	3.1
Rajasthan	3.1
Uttarakhand	2.9
Jammu and Kashmir	2.8
Nagaland	2.6
Madhya Pradesh	2.5
Tripura	2.5
Maharashtra	2.5
Odisha	2.4
Delhi	2.3
Gujarat	2.2
Jharkhand	2.2
Karnataka	2.1
Chhattisgarh	1.9
Goa	1.8
Punjab	1.8
Haryana	1.7
Kerela	1.7
Tamil Nadu	1.6
Lakshadweep	1.4
Telangana	1.4
Chandigarh	1.4
Himachal Pradesh	1.4
Andhra Pradesh	1.1
Dadra and Nagar Haveli and Daman & Diu*	0.8
Pondicherry	0.4
Sikkim	0.0
A&N Islands	-0.1
Arunachal Pradesh	-0.6
Total All-India	2.7

Notes: **This table is equally worrisome because there has been a surge in the issue of Aadhar cards in some states.**
Highlighted states are those which have a surge of over 2%. Since India's annual birth rate is 1.65%, any surge within 15 months of over 2% is curious, at best.
Interestingly, the largest increases are in border states, close to other populated countries.

Sources: **Govt of India, Lok Sabha**

(1) Unstarred question No. 1230 answered on 9th February 2022 on "Process of issuing Aadhar cards"
(2) Unique Identification Authority of India and unstarred question no 5519 answered on 5th April 2023 on "Aadhar of deceased persons".

Both replies were given by Rajeev Chandrashekhar, Minister of State for Electronics, and Information Technology.

Table 2: States with more Aadhar Cards (than their respective populations)

States and Union Territories	Saturation (%) Live
Lakshadweep	108.8
Delhi	108.7
Kerela	105.1
Himachal Pradesh	104.6
Goa	103.8
Punjab	102.6
Telangana	101.8
Haryana	101.2
Uttarakhand	100.9
Odisha	99.9
Andhra Pradesh	98.2
West Bengal	97.7
Mizoram	97.4
Tamil Nadu	97.1
A & N Island	97.0
Karnataka	96.5
Chandigarh	95.6
Puducherry	95.1
Maharashtra	94.9
Chhattisgarh	94.8
Rajasthan	94.0
Dadra and Nagar Haveli and Daman & Diu *	93.6
Uttar Pradesh	93.5
Tripura	93.2
Gujarat	92.8
Jharkhand	92.2
Madhya Pradesh	91.2
Bihar	87.5
Jammu and Kashmir	86.9
Assam	86.8
Sikkim	84.5
Manipur	82.9
Ladakh	80.3
Arunachal Pradesh	80.3
Meghalaya	69.4
Nagaland	61.6
Total All-India	94.8

Notes: Saturation is the percentage of Aadhar card holders to the population of the State. Greater than 100% means more Aadhar cards than the population. An urgent verification by the Census is required here.

Sources: Govt of India, Lok Sabha

(1) Unstarred question No. 1230 answered on 9th February 2022 on "Process of issuing Aadhar cards"
(2) Unique Identification Authority of India and unstarred question no 5519 answered on 5th April 2023 on "Aadhar of deceased persons".

Both replies were given by Rajeev Chandrashekhar, Minister of State for Electronics, and Information Technology.

updates at physical centers still incur a ₹50 fee, prompting concerns that this could be viewed as a revenue-generating measure.

The introduction of Article 139AA in the 2017 Finance Bill linked Aadhaar with PAN cards, resulting in a dramatic surge in PAN card issuances. Data from Dr. Anupam Saraph under RTI revealed that by March 2017, 82% of PAN cards were issued with Aadhaar, compared to just 10% before the bill, potentially opening the door for fraudulent access to funds. Further, inflated slum household numbers indicate that politically influential areas might be receiving up to ten times the subsidies they should, draining public resources. According to Asia Converge, slum households, with an average of just 0.5 persons per household compared to the national average of 5.3, maybe over-allocated subsidies. This surge in Aadhaar card issuance also inflates election rolls, escalating taxpayer burdens and fostering fraudulent benefit claims. Tables 1 and 2 underscore these discrepancies, shedding light on the vulnerabilities in the system's administration (Bhaskar, 2023).

b) Aadhaar's Implementation: Barriers and Exclusion

Aadhaar, India's biometric identification system, was launched with the ambitious goal of providing universal access to identification. However, its implementation has drawn criticism for marginalizing vulnerable groups. Despite the provisions of the Aadhaar Act of 2016, systemic flaws and exclusionary practices continue to deny access to essential services for many. As Raje and Pandey (2023) underscore, homeless individuals remain excluded because they often lack necessary documents like birth certificates. Respondents from this community shared how they found themselves entirely locked out of the Aadhaar system, unable to overcome bureaucratic hurdles.

Systemic Barriers and Marginalized Identities — Exclusionary practices in Aadhaar extend beyond the homeless population. Linking Aadhaar to PAN cards has exposed significant systemic gaps, particularly affecting transgender individuals. These individuals face discrimination due to the binary gender options — "male" and "female" — on official forms. According to a 2019 Dalberg study, 30% of homeless individuals and 27% of transgender people are without Aadhaar. Additionally, biometric challenges further hinder access. Workers in labor-intensive fields like sanitation reported that worn fingerprints often result in biometric identification failures, exacerbating their exclusion (Totapally et al., 2019).

One respondent from a religious minority described how clerical errors, such as incorrect name entries or mismatched age data on Aadhaar cards, led to months of delays and frustration. Ranita, a slum dweller, recounted her inability to access food benefits because her Aadhaar was not linked to her ration card, despite multiple visits to government offices. Another Muslim woman expressed her distress over a discrepancy in her surname: while her family title was "Sheikh," her Aadhaar card listed it as "Sk." This mismatch caused significant hurdles, preventing her from receiving the *Lakshmir Bhandar* monthly monetary aid from the government.

Food Security and the Digital Divide

Aadhaar's impact on food security has been particularly alarming. Technical glitches and identity verification failures have caused ration denials and, in extreme cases, starvation deaths. The Right to Food Initiative has highlighted that marginalized groups — such as Dalits, Muslims, and indigenous tribes — are disproportionately affected (Ganesh, 2018). The digital divide adds another layer of exclusion in rural areas, where poor internet connectivity and frequent cyber issues prevent residents from registering or updating their Aadhaar information. Respondents from remote villages explained how these barriers left them unable to access services promised as universally available.

Daily Life Disruptions

For many, Aadhaar has disrupted daily life. Rabi Ghosh, a garbage collector, faced immense difficulties after losing his Aadhaar card. Despite repeated visits to government offices, he found no resolution, losing valuable time and resources in the process. Similarly, Dinesh Sharma, a carpenter, voiced his frustration with the Aadhaar system, calling it unnecessarily complex compared to the straightforward voter ID card. With biting sarcasm, he said, "Are we supposed to turn into full-time Aadhaar trackers now? Every few months, there's some new requirement — link this, update that — just to prove our existence and claim the benefits we're entitled to. We're daily wage earners, not officegoers with paid leave. Who's going to pay for the day's wages we lose standing in endless queues at these camps?"

Elderly individuals have faced unique challenges. Malati Rani De, a 70-year-old woman, was unable to access essential services because her Aadhaar card mistakenly listed her father's name instead of her husband's. Rectifying this seemingly minor clerical error required multiple visits to government offices, causing significant delays and frustration. Like Malati, many senior citizens struggle with Aadhaar's technological processes, often relying on younger family members for assistance. This dependency has left them feeling disempowered and vulnerable.

Biometric Challenges and Economic Exploitation

Aadhaar's reliance on biometric data for identification and financial transactions has introduced further complications. The government's emphasis on biometrics, encapsulated by the slogan "Your thumbprint is your bank," overlooks the vulnerabilities of this system. Manual laborers, whose fingerprints are often worn, and individuals with medical conditions, such as cataracts, frequently encounter biometric failures. For example, one respondent shared how cataracts rendered iris scans ineffective, leading to a two-year denial of services (Sobti, Sahni, Bala, 2020). As one resident observed, Aadhaar's data-driven governance has become a barrier rather than a bridge to essential services.

These challenges are exacerbated by the difficulty of correcting errors within the Aadhaar system. An ID Insight report found that Aadhaar's demographic error rate is significantly higher than that of voter IDs (Abraham et al., 2018). A migrant laborer shared how his family was denied rations due to technical errors, highlighting the flawed nature of a system meant to ensure equity (Right to Food Initiative).

Adding to the burden, instances of economic exploitation have surfaced. Respondents reported being charged unofficial fees of up to ₱500 for Aadhaar registration or updates, disproportionately impacting economically weaker sections. Reports of students in Mumbai manipulating biometric attendance by reproducing fingerprints further reveal vulnerabilities in the system (Shiva and Madan, 2020).

An economics professor quipped that Aadhaar, hailed as the great equalizer, seems to have mastered the art of selective exclusion instead. Designed to foster equity, it ironically excels at deepening social divides, leaving the homeless, religious minorities, transgender individuals, and the elderly grappling with biometric failures, clerical errors, and endless bureaucratic mazes. "Instead of bridging gaps," they remarked wryly, "Aadhaar has created an Olympic sport out of survival — jump through enough hoops, and maybe, just maybe, you'll get what's already yours. But hey, who needs dignity when you have a 12-digit number?"

(c) Surveillance State Concerns: Aadhaar's Social Impact

The earlier section takes a ground-level view, focusing on practical hurdles that make Aadhaar a logistical barrier to essential services like food and financial aid. It highlights issues such as biometric failures, documentation gaps, and technical glitches that disproportionately impact marginalized communities. In contrast, this section examines the broader societal risks tied to Aadhaar. It critiques the system as a potential tool for authoritarian control, capable of reinforcing systemic discrimination, fostering religious bias, and eroding privacy and civil liberties. While one explores the day-to-day struggles of citizens, the other warns of the larger implications for governance and social equity. Together, they reveal a system grappling with immediate flaws and far-reaching dangers.

Aadhaar, envisioned as a tool for inclusivity, instead lays bare a paradox of inclusion. On one hand, it aims to streamline governance and expand access to essential services. On the other, its systemic flaws and coercive implementation disproportionately alienate the very communities it seeks to empower. Combining individual narratives with a broader critique reveals a system that struggles with both logistical failures and governance concerns.

Ground-Level Struggles

Aadhaar's systemic inefficiencies — biometric failures, clerical errors, and technical glitches — create a logistical maze for marginalized communities, blocking access to essential services. Rehana Sheikh, a vendor, endured months of hunger due to a spelling error on her Aadhaar card that delayed her benefits. Dinesh Sharma, a carpenter, criticized the endless mandates: *"Do we spend our lives proving we exist?"* Ranita Kumari, a slum

dweller, faced food aid denial over a missing Aadhaar link, while Ravi Ghosh, a garbage collector, languished in bureaucratic limbo after losing his card. Elderly individuals like Aarati Devi spoke of the humiliation of relying on intermediaries to correct errors: *"Each trip costs time, money, and dignity."* These stories highlight how Aadhaar often exacerbates hardships rather than alleviating them.

Zooming Out: Societal Risks

Critics argue that Aadhaar's flaws go beyond logistical barriers, cautioning against its potential as a tool for authoritarian control. Embedded in India's digital framework, Aadhaar risks fostering a surveillance state, critics contend, enabling systemic discrimination, religious bias, and privacy violations (Ganesh, 2018; Khera, 2018; Browne, 2015).

Aslam Qureshi, a daily laborer, voiced fears of Aadhaar misuse: *"I hear minority accounts can be frozen over 'mismatches.' Without Aadhaar, other IDs mean nothing."* Such concerns are worsened by abrupt mandates and conflicting directives, leaving citizens alienated. Rural residents also reported how technical glitches and data errors eroded their trust in the system.

Recent incidents have heightened these anxieties. During the 2024 Kanwar Yatra, reports emerged of dhaba owners being told to display the Aadhaar cards of Hindu employees while dismissing Muslim staff. Such actions stoke fears of Aadhaar institutionalizing religious discrimination, further fracturing social harmony.

Aadhaar: Facilitator or Enforcer?

Despite assurances that Aadhaar isn't mandatory for services like bank accounts, citizens frequently face enforced adherence. One exasperated customer explained: *"I provided my passport, voter ID, and PAN, but the bank insisted on Aadhaar. Even online applications were rejected due to a so-called mismatch."* These contradictions between official policies and real-world practices reveal how Aadhaar's implementation often coerces compliance, disproportionately impacting those with limited resources.

Justice Chandrachud has critiqued Aadhaar's premise, asserting that *"dignity to the marginalized cannot override the right of a person to bodily autonomy."* Khera (2018) recounts the tragic death of an 11-year-old girl removed from the subsidy register because her Aadhaar wasn't linked to her ration card. For many elderly and disabled individuals, Aadhaar has added hurdles rather than alleviating them, requiring physical presence and leaving them vulnerable to exploitation by intermediaries.

Thus, Aadhaar, hailed as a tool for inclusion, often deepens societal inequities. Supporters claim it combats fraud and streamlines services, but its coercive implementation and systemic flaws — ranging from administrative inefficiencies to fears of surveillance — alienate marginalized groups. Rather than fostering equity, Aadhaar exacerbates the digital divide, leaving the vulnerable further behind and its promise of inclusivity largely unfulfilled.

(d) The Dark Side of Aadhaar: Security Breaches and Exploits

The Aadhaar system, envisioned as a cornerstone of Digital India, faces mounting scrutiny over its alarming vulnerability to misuse and breaches. Reports suggest that the dark web, notorious for illegal activities, is increasingly trafficking Aadhaar-related data, heightening concerns about security and potential exploitation. The *Economic Times* highlights a surge in Aadhaar-linked banking fraud, urging individuals to lock their biometrics as cybercriminals exploit the Aadhaar-enabled Payment System (AePS), causing significant financial losses (*The Economic Times*, 2023). Civil society groups like Bank Bachao Desh Bachao Manch have warned the Reserve Bank of India about banks coercing account holders into linking Aadhaar without consent, leaving them exposed to fraud despite UIDAI's guidelines (Kulkarni, 2023).

High-profile incidents underline the systemic flaws. In 2018, a government website accidentally published residents' Aadhaar details, including names, identity numbers, and photographs, which were swiftly taken down (Hameed, 2023). Similarly, the Andhra Pradesh government exposed over 130,000 Aadhaar numbers, some with bank details, in another grave breach (Goel, 2018). Instances of Aadhaar misuse continue to grow, such as fraudsters in Chandigarh using fake Aadhaar cards to purchase high-end phones (Hameed, 2023) or criminals in Gurgaon leveraging leaked Aadhaar data to implicate an innocent engineer in false allegations (Hameed, 2023). In a broader context, government websites and departments have repeatedly shared sensitive Aadhaar-linked data online, violating confidentiality norms (Dixit, 2017a; Goel, 2018).

The gravity of the situation was further exposed when *The Tribune* revealed in 2018 that personal UIDAI information could be purchased for a mere Rs. 500, demonstrating how easily the system can be compromised (Khaira, 2018). More recently, *The Economic Times* reported in October 2023 on one of India's largest data breaches, where a staggering 81.5 crore personal records, including Aadhaar and passport information, were found for sale on the dark web. The leaked data, reportedly shared by a threat actor, contained samples of 100,000 records (*The Economic Times, 2023*). Adding to the crisis, Rajasthan authorities uncovered a major scam involving forged Aadhaar cards and biometric data, including that of children, sold for Rs. 25,000 (*The Economic Times, 2024*).

Despite these troubling patterns, UIDAI has downplayed concerns, claiming that Aadhaar data breaches have not occurred and citing robust encryption and authentication safeguards for biometric data (Balaji, 2017). However, investigations and reports paint a different picture, exposing loopholes that embolden misuse. Sindhuja Balaji of *Forbes Asia* notes that the government and private firms, such as AuthBridge and Signzy, are working on initiatives like India Stack and FIDO-compliant biometric technologies to strengthen Aadhaar's security (Balaji, 2017). Yet, these efforts remain outpaced by the scale of exploitation, underscoring the urgent need for systemic reform.

The recurring breaches and frauds surrounding Aadhaar reveal a precarious reality: while its intent is to empower citizens and drive digital inclusion, its vulnerabilities pose serious risks to privacy and security. Without comprehensive safeguards and account-

ability mechanisms, Aadhaar risks becoming a liability rather than an enabler of trust and progress in India's digital transformation.

(e) Resistance and Reform: Towards a Balanced Framework

The Aadhaar system's vulnerabilities have not only exposed individuals to privacy and security risks but also sparked debates over government surveillance and the scope of individual rights. These concerns have led to landmark judicial interventions, shaping the framework of Aadhaar's use while highlighting systemic shortcomings.

Between 2013 and 2015, the Supreme Court of India consistently maintained that Aadhaar should remain voluntary, reflecting opposition to its coercive implementation (Ghoshal, 2017). However, a turning point came in August 2017, when the Court recognized privacy as a fundamental right under Article 21 of the Indian Constitution. This ruling underscored privacy's broad protection and challenged Aadhaar's expansive data collection. Despite this judgment, the practical enforcement of privacy protections has been inconsistent. For example, Aadhaar was recently removed as a mandatory requirement for accessing food rations in Delhi, but similar reliefs have not been uniformly applied (Goel, 2018).

The cause-and-effect relationship between Aadhaar's implementation and judicial scrutiny is best illustrated in the 2017 landmark case *Justice K. S. Puttaswamy (Retd.) & Anr. vs. Union of India & Ors* (SC Observer, 2017). Here, the Supreme Court unequivocally declared privacy as a fundamental right, overruling earlier judgments that failed to recognize this principle. This declaration placed a constitutional check on Aadhaar's practices, questioning the legitimacy of its biometric data collection.

The 2018 case *Puttaswamy (Aadhaar-5J.) v. Union of India* further explored the balance between Aadhaar's utility and individual rights. The Court upheld the scheme's constitutionality but imposed critical limitations. It ruled that Aadhaar could only be mandatory for government welfare benefits, not for bank accounts, mobile connections, or school admissions (Supreme Court of India, 2021). Moreover, the judgment emphasized the pressing need for stronger data protection and security measures to mitigate Aadhaar's risks.

These judicial pronouncements highlight the tension between Aadhaar's objectives and the need for a privacy-first approach. While the Court sought to curtail Aadhaar's overreach, systemic changes to its implementation remain uneven. This interplay between Aadhaar's vulnerabilities, judicial oversight, and government policies continues to shape the trajectory of digital governance in India.

(f) Conflicts of Interest and Consent Issues in Aadhaar's Modern Framework

The involvement of private companies like OnGrid and Khosla Labs in Aadhaar operations raises concerns about the privatization of sensitive public services. These companies handle e-KYC verification and digital identity infrastructure, but their deep integration

with Aadhaar creates potential conflicts of interest, particularly due to ties with UIDAI (Thaker, 2018). Users are often unaware of the implications of sharing Aadhaar-linked data, especially when alternatives are limited.

The "revolving door" between public and private sectors intensifies these concerns, with former UIDAI officials potentially gaining early access to sensitive data, raising ethical issues about privacy and public trust. Security vulnerabilities have been exposed, such as The Tribune's report on Aadhaar data being accessed for a fee (Bansal, 2018), and Airtel's misuse of Aadhaar-based e-KYC for unauthorized accounts. While UIDAI has introduced stricter regulations, gaps persist, especially for marginalized groups who face coerced consent.

Companies like OnGrid, TrustID, and IDfy, involved in Aadhaar-based background checks, and the movement of UIDAI officials to the private sector, further exacerbate conflicts of interest. In 2018, financial tech firms were barred from accessing Aadhaar data amid legal challenges, and iSpirt, which manages India Stack, has been criticized for its ties to former UIDAI officials and global corporations like Microsoft and Facebook (Dharmakumar, 2018).

Additionally, the expansion of State Resident Data Hubs (SRDHs) raises privacy concerns, as these hubs store extensive personal data. Vulnerabilities in SRDH systems have been exposed (Baptiste, 2018), and Aadhaar's use for creating centralized citizen profiles in states like Andhra Pradesh raises fears of mass data collection. The Telangana Police's use of Aadhaar to profile offenders illustrates the risks associated with excessive data collection (Sethi, 2018). The concept of a 360-degree database, where extensive personal data is stored with minimal oversight, violates the Aadhaar Act's privacy principles, posing significant risks to personal data security (Venkatanarayanan, 2017).

(g) Privacy Frameworks: A Tale of Three Systems

India's Aadhaar system, hailed as the largest biometric identification project in the world, was introduced to streamline service delivery and reduce fraud. However, despite its lofty promises, Aadhaar operates in a vacuum of robust data protection laws, leaving individuals vulnerable to privacy violations. Justice D. Y. Chandrachud, during the landmark Supreme Court hearings on Aadhaar, voiced strong concerns, calling the project "completely violative of privacy." He warned against reducing constitutional guarantees to mere experiments with "probability algorithms and technological vicissitudes" (Sen, 2024). Similarly, economist Jean Drèze critiqued the Unique Identification Authority of India (UIDAI) for its data practices, alleging that Aadhaar data is accessible to anyone willing to pay, with only core biometrics offered minimal protection (Sen, 2024).

In contrast, the General Data Protection Regulation (GDPR) of the European Union and the U. S.'s Social Security Number (SSN) system illustrate how strong regulatory frameworks, and targeted oversight can bolster data security and privacy — areas where Aadhaar remains vulnerable.

The Loopholes of Aadhaar

Aadhaar's legal backbone, the Aadhaar Act, 2016, was heavily criticized for bypassing proper legislative procedures, being enacted as a Money Bill. Justice Chandrachud deemed it unconstitutional and critiqued its failure to adequately safeguard privacy, dignity, and autonomy (Supreme Court Observer, 2018). Subsequent attempts to regulate Aadhaar through the Personal Data Protection Bill (2019, 2023) and the Digital Personal Data Protection Bill, 2023, have been marred by glaring deficiencies. These laws permit government agencies to collect and retain data indefinitely, bypassing consent under vague clauses. Provisions such as Section 36 allow the government to compel private companies to share data, while Clause 3(c)(ii) authorizes the processing of publicly available data without consent. These gaps exacerbate the risk of misuse, especially in a centralized system like Aadhaar, where massive amounts of biometric and demographic data are stored.

The system's emphasis on implementation over security has led to a litany of breaches. Critics argue that Aadhaar lacks mechanisms like data portability, the right to be forgotten, or rigorous safeguards for cross-border data transfers, unlike its Western counterparts (Crawford et al., 2023). Instead of fortifying individual rights, Aadhaar's governance expands state power, as demonstrated by its expansive mandates for linking to essential services, which further alienate marginalized groups.

GDPR: A Gold Standard for Privacy

In stark contrast, the General Data Protection Regulation (GDPR) has set a global benchmark for privacy protection since its enactment in 2018. Rooted in principles like privacy by design and transparency, GDPR enforces stringent consent requirements — consent must be explicit, informed, and revocable, particularly for sensitive data. It also grants individuals robust rights, such as the ability to access, correct, erase, and even port their data. Organizations must implement measures like encryption and anonymization, ensuring data security at every step (Hoofnagle, van der Sloot, Borgesius, 2019).

GDPR's enforcement framework is equally formidable. Violators face fines up to €20 million or 4% of global annual turnover, creating powerful incentives for compliance. Enforcement actions by the EU have targeted tech giants like Meta and Google, with penalties reflecting GDPR's focus on transparency and user rights. For instance, Meta Ireland was fined €1.2 billion in 2023 for transferring EU user data to the U. S. without adequate safeguards, while Google was penalized €50 million in 2019 for opaque data practices in advertising (CNIL, 2019-2023). These cases illustrate GDPR's emphasis on accountability and its transformative impact on global data governance.

The U. S. SSN: A Patchwork Framework

The U. S. Social Security Number (SSN) system operates under a patchwork of sector-specific laws. While it lacks the unified rigor of GDPR, oversight mechanisms like the

Next Generation Identification (NGI) system offer better biometric data management than Aadhaar. The NGI incorporates checks against unauthorized use, mitigating risks like identity theft. However, the SSN system's fragmented legal framework leaves gaps in consent requirements and individual control, exposing users to data misuse. Despite these shortcomings, the decentralized nature of SSN data reduces the risk of large-scale breaches, a vulnerability inherent to Aadhaar's centralized database (Dixon, 2017).

Aadhaar, though a bold vision for digital identity, falls short due to weak privacy protections. Unlike GDPR, which prioritizes user rights with strict enforcement, Aadhaar's framework leans towards state interests, undermining individual autonomy. Its centralized structure further heightens vulnerabilities compared to the decentralized U.S. SSN system. To realize its potential, Aadhaar must adopt a privacy-first approach, with robust laws that protect individuals and ensure accountability. Until then, it remains an ambitious but flawed initiative, lagging global standards.

Discussion

(a) Aadhaar: Bridging Modernity and Tradition or Reinforcing Inequities in India's Identity Landscape?

Aadhaar, India's AI-driven biometric identity system, is a vivid example of how global technological frameworks intersect with the unique socio-cultural fabric of India. By drawing on Homi Bhabha's theories of hybridity and mimicry, Aadhaar reflects the complexities of merging universal ideals into a diverse and deeply traditional society. In doing so, it reshapes individual and collective identities in a post-colonial context. Homi Bhabha's concept of hybridity illuminates Aadhaar's dual role in combining global principles, such as fostering inclusion by connecting citizens to government services, with India's intricate realities shaped by caste, ethnicity, and regional diversity. This duality is marked by a paradox: while Aadhaar aspires to inclusivity, it often ends up excluding marginalized communities, reinforcing the very social inequalities it aims to bridge. Issues like biometric authentication failures and data inaccuracies disproportionately affect those already on the fringes of society, highlighting the friction between technological modernity and deeply rooted social traditions (Ramanathan, 2018). This dynamic reflects what Bhabha (1994) describes as the "interstitial passage," where old and new identities coexist in an uneasy balance. For millions, navigating the Aadhaar system requires reconciling personal and cultural identities with the state-imposed definitions of a digital self. This tension creates a mechanism that promises empowerment but often delivers alienation, as individuals struggle to find equilibrium between their lived realities and the demands of a standardized system. By promising equality while perpetuating disparities, Aadhaar embodies the inherent contradictions of post-colonial governance.

Bhabha's theory of mimicry offers another lens through which to understand Aadhaar's complexities. Mimicking global biometric systems, Aadhaar is "almost the same, but not quite," as Bhabha (Bhabha, 1994: 122) puts it. While it strives to emulate robust

frameworks like the European Union's GDPR, it falls short in critical areas, particularly around privacy and data protection. The involvement of private entities, such as OnGrid and Khosla Labs, in managing sensitive public data raises significant ethical concerns (Ganesh, 2018). These challenges are compounded by operational inefficiencies, including instances of duplicate Aadhaar records and errors caused by migration or demographic changes. Together, these flaws highlight the limitations of imposing a universal framework in a society as diverse and stratified as India's. The challenges of Aadhaar extend beyond technical and operational shortcomings. Its algorithm-driven identity system often clashes with India's traditional markers of recognition, such as community ties and storytelling traditions. These culturally rooted forms of identity offer a sense of belonging that Aadhaar's standardized, digital approach cannot replicate (Kshirsagar, 2024). For many, this disconnect represents a loss of connection to their cultural heritage, further emphasizing the difficulties of harmonizing global governance systems with India's deeply ingrained social structures.

The risks associated with Aadhaar's implementation are not merely theoretical. Its expansive data collection and recurring security breaches evoke unsettling parallels with colonial practices. Historically, tools like censuses and anthropological records were used by colonial regimes not to empower but to control and classify populations (Said, 1978). Simone Browne (2015) critiques contemporary surveillance technologies, highlighting how these systems often reinforce social hierarchies rather than dismantle them. Aadhaar's vulnerabilities — evidenced by cases of data being leaked onto the dark web — illustrate how such systems can perpetuate rather than mitigate power imbalances. Far from empowering citizens, these breaches increase their exposure to risks, reinforcing a troubling legacy of dominance disguised as governance.

At its core, Aadhaar exemplifies the contradictions of post-colonial modernity. While it aspires to inclusivity and efficiency, its implementation risks alienating the communities it seeks to uplift. For Aadhaar to truly fulfill its promise, it must move beyond its current framework of standardization and surveillance. By embracing cultural sensitivity, addressing privacy concerns, and safeguarding individual rights, Aadhaar can evolve into a system that genuinely empowers India's diverse population. Currently, Aadhaar remains a contested symbol of identity in the digital age. It stands at the crossroads of tradition and modernity, inclusion and exclusion, empowerment and control. The system's success will ultimately depend on its ability to reconcile these tensions, transforming from a source of alienation into a bridge between global aspirations and local realities.

b) Surveillance, Identity, and Power: Aadhaar as a Modern Panopticon in Post-Colonial India

Aadhaar embodies Michel Foucault's concept of the Panopticon, where pervasive surveillance compels individuals to regulate their own behavior. By requiring biometric authentication for services such as banking, welfare, and taxation, Aadhaar establishes conformity to state-defined norms. Individuals internalize this compliance to avoid ex-

clusion, turning participation into a prerequisite for citizenship (Galic, Timan, Koops, 2016). The system subtly coerces alignment with its biometric framework, reinforcing state authority through penalties for noncompliance (Sundquist, 2023). Situated between tradition and modernity, Aadhaar reflects the complexities of post-colonial governance, where identity is both empowering and controlling.

Drawing on Foucault's theories of disciplinary power and biopower, Aadhaar echoes colonial surveillance and identity regulation mechanisms. Much like historical censuses and anthropological classifications, it reduces individuals to data points recognized by the state, marginalizing those who fail to fit into rigid frameworks. Its biometric foundation creates a paradox: while promising modernization, it perpetuates exclusion and deepens societal hierarchies (Ganesh, 2018). Aadhaar exemplifies biopower by digitizing identity, controlling it through the state, and marginalizing those who do not conform. Vulnerable groups such as the homeless or transgender individuals face exclusion due to mismatched biometrics or lack of documentation, perpetuating social inequities (Beydoun, 2018; Cisney, Morar, 2016). Sundquist (2023) underscores how pervasive surveillance technologies like Aadhaar become normalized, perceived as an inevitable trade-off for convenience.

Aadhaar's design reflects Foucault's concept of power as a force that operates through normalization and surveillance. Biometric verification for accessing essential services embeds state oversight into everyday life. The periodic updating of Aadhaar details deepens reliance on bureaucratic structures, reinforcing conformity to national norms. This mirrors Foucault's notion of "docile bodies," where individuals internalize societal expectations and become compliant participants in their own regulation (Foucault, 1991). Through these mechanisms, Aadhaar aligns individual identity with a state-imposed framework, diminishing the autonomy of personal and cultural self-definition.

The expansive scope of Aadhaar — covering services from LPG subsidies to healthcare, ration distribution, and tax filing — entangles citizens in a continuous cycle of registration and re-registration. While the system aims to integrate the unregistered into formal systems, it risks excluding those who cannot meet its stringent requirements. Failure to authenticate biometrically or provide the necessary documentation results in denial of critical services, effectively rendering parts of the population invisible to the state. This exclusion reinforces power imbalances and mirrors colonial identity management strategies, where bureaucratic structures served as tools for population control.

Aadhaar also exemplifies Foucault's insight into the role of surveillance in constructing identities. By reducing identity to biometric markers, it prioritizes digital conformity over individuality, marginalizing those who resist or fail to conform to its rigid frameworks (Foucault in Rabinow, 1991). The reliance on surveillance for identity validation entrenches exclusion within India's digital infrastructure, reinforcing social inequalities rather than dismantling them. As Browne (2015) observes, surveillance technologies often sustain social boundaries, embedding control within the governance structure.

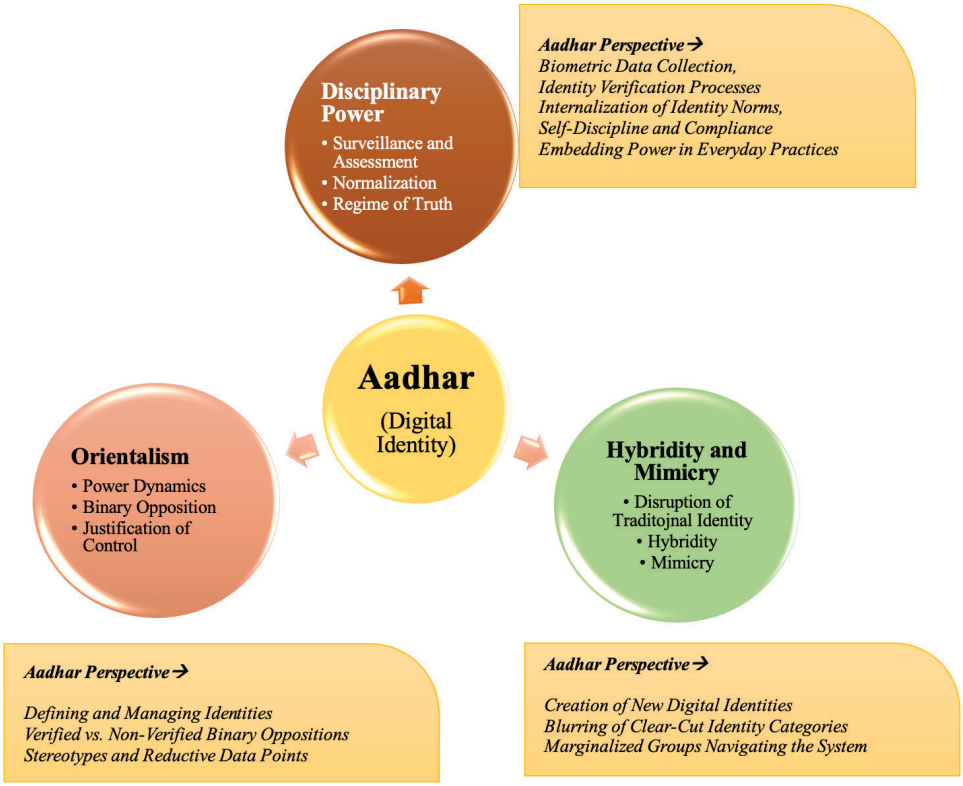
Ultimately, Aadhaar highlights the contradictions inherent in modern identity systems within a post-colonial society. While it aims to promote inclusion and empowerment, it often exacerbates exclusion and deepens historical inequities. By embedding global ideals of efficiency and modernization into India's complex social and cultural fabric, Aadhaar symbolizes contested identities, navigating the tensions between empowerment and surveillance, visibility and invisibility. It serves as a reminder of the delicate balance between the promises and dangers of a digital identity framework, raising questions about whether such systems can truly serve a diverse and unequal society.

(c) Aadhaar and the Legacy of Colonial Control: A Post-Colonial Analysis through Said's Lens

Aadhaar, India's vast biometric identity system, aims to streamline governance but mirrors colonial practices of control and categorization. Drawing on Edward Said's concept of Orientalism, Aadhaar reflects how post-colonial societies, in their pursuit of Western efficiency, may inadvertently internalize exclusionary frameworks rooted in colonial power structures (Al-Yusufi, 2003). Said's critique highlights how colonial regimes used tools such as maps, censuses, and anthropological studies not only for documentation but as instruments of domination and control. Similarly, Aadhaar's reliance on biometric data reduces individuals to state-sanctioned data points, reviving historical practices of surveillance and classification.

Though designed to provide a universal identity for India's diverse population, Aadhaar risks imposing rigid categories that echo colonial governance strategies. As Said (1978) noted, colonial systems often restricted individuals to fixed, oversimplified identities, limiting opportunities for self-definition. Aadhaar exemplifies this paradox, blending modern governance with enduring colonial legacies. Its push for a universal digital identity threatens to erase India's rich cultural diversity, where identity is traditionally shaped by factors such as caste, region, and community.

Edward Said's concept of cultural homelessness is relevant to Aadhaar's framework, as it imposes a singular, state-defined identity that marginalizes those who do not fit within its narrow parameters. Individuals from diverse backgrounds — transgender persons, informal laborers, government officials, or entrepreneurs — cannot be reduced to a single, homogenized identity. The system thus creates a form of exile, alienating people from their unique identities by forcing conformity to an external, rigid classification (Said, 1978). Aadhaar's power also invites resistance, as Foucault (1998) suggested. Legal challenges and critiques have reframed Aadhaar as a tool of surveillance and exclusion, advocating for a more inclusive approach to identity (PowerCube, 2023).



The discussion is summarized through a conceptual map that illustrates the relationship between Aadhaar, digital identification systems, and the frameworks of postcolonial and post-structuralist theory. This map explores how Aadhaar, as a digital identification tool, intersects with themes of power, surveillance, and governance as analyzed by post-structuralist

The discussion is summarized through a conceptual map that illustrates the relationship between Aadhaar, digital identification systems, and postcolonial and post-structuralist theory. This map examines how Aadhaar, as a digital identification tool, intersects with themes of power, surveillance, and governance, drawing on post-structuralist analysis. It explores how the system reinforces or disrupts colonial legacies of control, surveillance, and exclusion while shaping social identities and interactions. Through this lens, the map highlights the complexities of Aadhaar as both a tool for governance and a mechanism that may perpetuate new forms of social stratification, reflecting broader theoretical concerns regarding biopolitics, power dynamics, and identity in postcolonial contexts.

Conclusion

This study examines how Aadhaar, India's biometric identity program, reshapes the relationship between citizens and the state, blending inclusion with surveillance. Beyond its role as a technological tool, Aadhaar becomes a mechanism of governance, discipline, and identity formation, with profound ethical and social implications. Empirical findings highlight systemic flaws, technical failures, and exclusionary practices that undermine Aadhaar's promise of inclusion. Data errors, inequities faced by marginalized groups, and privacy vulnerabilities expose the paradox of a system designed to universalize access while deepening societal divides.

Through post-colonial and post-structuralist perspectives, Aadhaar is analyzed as a continuation of colonial classification and control, yet it aspires to modernity. Foucault's concepts of disciplinary power and biopolitics reveal how Aadhaar normalizes compliance, transforming identity into a state-regulated digital construct. Said's theory of cultural dominance positions Aadhaar as a homogenizing force, threatening cultural diversity. Bhabha's idea of hybridity provides a lens for resistance and adaptation, suggesting that Aadhaar could be reimagined as a tool for equitable governance.

The study advocates for stronger regulatory frameworks, inspired by models like the GDPR, to protect privacy and human rights. It calls for redesigning Aadhaar to consider the realities of vulnerable populations, ensuring that technological governance promotes dignity, equity, and justice. Ultimately, Aadhaar reflects India's values and power dynamics, and its future impact will depend on the decisions made by policymakers, civil society, and citizens.

Future research could involve a longitudinal study over the next few years to assess Aadhaar's normalization as a governance tool and evaluate corrective measures by the government. Expanding the study to include diverse states, rural-urban sectors, and various socio-economic classes would offer a broader perspective. Incorporating empirical fieldwork and case studies, particularly of marginalized groups like transgender individuals, would provide deeper insights into Aadhaar's evolving impact on identity, privacy, and governance.

References

Agrawal S., Banerjee S., Sharma S. (2017) Privacy and security of Aadhaar: A computer science perspective. *Economic and Political Weekly,* vol. 52, no 37, pp. 93-102.

Ajana B. (2010) Recombinant identities: Biometrics and narrative bioethics. *Journal of Bioethical Inquiry,* vol. 7, no 2, pp. 237–258.

Andreotti V. D. O. (2011) Homi Bhabha's contribution and critics. *Actionable postcolonial theory in education,* Palgrave Macmillan, pp. 25-35.

Ashcroft B., Aluhwalia P. (1991) *The post-colonial studies reader,* Routledge.

Bansal S. (2018, September 27) Supreme Court judgment on Aadhaar leads to confusion in private sector. *Hindustan Times*. Retrieved from https://www.hindustantimes.com/

india-news/supreme-court-judgment-on-aadhaar-leads-to-confusion-in-private-sector/story-1RLacutrkA4EU4Zx2iVlUK.html

Basu S., Omotuboraa A. (2024, September 9) Challenging the coloniality in global AI regulation frameworks. *DNLU Student Law Journal*. Retrieved from https://dnluslj.in/challenging-the-coloniality-in-global-ai-regulation-frameworks/

Beydoun K. (2018) *Surveillance and the surveillance state: The impact on marginalized communities*, Oxford University Press.

Beydoun K. A. (2018) Biometrics and marginalized populations. *Human Rights Quarterly*.

Bhabha H. (1984) Of mimicry and man: The ambivalence of colonial discourse. *October,* no 28, pp. 125–133.

Bhabha H. (1986) Difference, discrimination, and the discourse of colonialism. *Literature, politics and theory. Papers from the Essex Conference, 1976-1984* (F. Barker et al. eds.), Methuen.

Bhabha H. K. (1986) Foreword. *The post-colonial studies reader.* (B. Ashcroft, G. Griffiths, H. Tiffin eds.), Routledge.

Bhabha H. K. (1994). *The location of culture,* Routledge.

Bhaskar R. N. (2023, April 28) India's election peril, Aadhaar and the missing Census: Bogus Aadhaar cards imperil both Census and elections. *Substack*. Retrieved from https://bhaskarr.substack.com/p/indias-election-peril-aadhaar-and

Boehmer E. (1995). *Colonial and postcolonial literature: Migrant metaphors* (Illustrated ed.), Oxford University Press.

Brindaalakshm, K. (2013, May 10). Indian Govt. approves ₹3436.16 Cr. funds for Aadhaar Project Phase IV. *Medianama*. Retrieved from http://www.medianama.com/2013/05/223-indian-govt-approves-rs-3436-16cr-funds-for-aadhaar-project-phase-iv/

Browne S. (2015) *Dark matters: On the surveillance of blackness,* Duke University Press.

Castells M. (2000) *The rise of the network society,* Wiley-Blackwell.

Cisney V. W., Morar N. (2016) *Biopower: Foucault and beyond*, University of Chicago Press.

Cisney V. E., Morar N. (2016). *Foucault and biopolitics: The governance of populations,* Routledge.

Crawford G., Maclean F., van der Merwe D., Burrell K., Lee B. H., Park A., Vasile I., Smyth A. (2023, December 13) India's Digital Personal Data Protection Act 2023 vs. the GDPR: A comparison. *Latham & Watkins*. Retrieved from https://www.globalprivacyblog.com/2023/12/indias-digital-personal-data-protection-act-2023-vs-the-gdpr-a-comparison/

Creswell J. W. (2009) *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.), Sage Publications Inc.

Daigle B., Khan M. (2020) The EU General Data Protection Regulation: An analysis of enforcement trends by EU data protection authorities. *Journal of International Commerce and Economics.* Retrieved from https://www.usitc.gov/journals

De Loo I., Lowe A. (2011) Mixed methods research: Don't — 'just do it'. *Qualitative Research in Accounting & Management,* vol. 8, no 1, pp. 22-38.

Deep A. (2024, January 26) Government makes citizenship disclaimer on new Aadhaar cards more prominent. *The Hindu*. Retrieved from https://www.thehindu.com/news/national/government-makes-citizenship-disclaimer-on-new-aadhaar-cards-more-prominent/article67780076.ece

Deep A. (2024, January 26) Government makes it clear that Aadhaar no proof of citizenship, birth date. *The Hindu*. https://www.thehindu.com/news/national/government-makes-citizenship-disclaimer-on-new-aadhaar-cards-more-prominent/article67780076.ece

Dixon P. (2017) A failure to "do no harm" — India's Aadhaar biometric ID program and its inability to protect privacy in relation to measures in Europe and the U. S. *Health Technology,* vol. 7, no 4, pp. 539-567.

Dreze J. (2010, November 25) Unique facility or recipe for trouble. *The Hindu*. Retrieved from http://www.thehindu.com/opinion/op-ed/unique-facility-or-recipe-for-trouble/article911055.ece

Easthope A. (1998) Homi Bhabha, hybridity and identity, or Derrida versus Lacan. *Hungarian Journal of English and American Studies (HJEAS),* vol. 4, no 1/2, pp. 145-151. Centre for Arts, Humanities and Sciences (CAHS), University of Debrecen.

Economic Times. (2023, October 31) Aadhaar data leak: Personal data of 81.5 crore Indians on sale on dark web: Report. *The Economic Times.* https://economictimes.indiatimes.com/tech/technology/aadhar-data-leak-personal-data-of-81-5-crore-indians-on-sale-on-dark-web-report/articleshow/104856898.cms

Economic Times. (2024, August 3) Aadhaar cards forged with toe prints, school kids' retina scans for Rs 25,000; CBI lodges FIR. *The Economic Times.* https://economictimes.indiatimes.com/news/india/aadhaar-cards-forged-with-toe-prints-school-kids-retina-scans-for-rs-25000-cbi-lodges-fir/articleshow/112250362.cms

Escobar A. (1995) *Encountering development: The making and unmaking of the third world,* Princeton University Press.

Fanon F. (1967) *Black skin, white masks,* Grove Press.

Foucault M. (1991) *Discipline and punish. The birth of the prison* (A. Sheridan, Trans.), Penguin.

Foucault M. (1998) *The history of sexuality. The will to knowledge* (R. Hurley, Trans.), Penguin.

Foucault M. (2003) *Society must be defended. Lectures at the Collège de France 1975–1976,* Picador.

Foucault M. (2007) *Security, territory, and population. Lectures at the Collège de France 1977–1978,* Palgrave.

Foucault M. (2009) *The birth of biopolitics. Lectures at the Collège de France 1978–1979,* Palgrave.

Galic M., Timan T., Koops B. J. (2016) Bentham, Deleuze, and beyond: An overview of surveillance theories from the panopticon to participation. *Philosophy & Technology,* vol. 30, no 1, pp. 9–10.

Galic M., Timan T., Koops B. J. (2016) *Surveillance and self-regulation: Theoretical foundations and practical implications*, Springer.

Ganesh, K. (2018) Aadhaar and the governance of identity in India. *Economic and Political Weekly,* vol. 53, no 3.

Ganesh K. (2018) The exclusionary impacts of Aadhaar: How technology marginalizes vulnerable groups. *Journal of South Asian Studies,* vol. 10, no 2, pp. 45-60.

Ganesh M. I. (2018, January 25) Data and discrimination: Fintech, biometrics, and identity in India. *Cyborgology.* Retrieved from https://thesocietypages.org/cyborgology/2018/01/25/fintech-aadhaar-and-identity-in-india/

Ganesh P. (2018) The risks of Aadhaar: Digital discrimination and privacy issues. *Indian Journal of Law and Technology.*

Gaventa J. (2003) Power after Lukes: An overview of theories of power since Lukes and their application to development (First draft). *Powercube.* Retrieved from https://www.powercube.net/wp-content/uploads/2009/11/power_after_lukes.pdf

Gaventa J. (2003) *Power and powerlessness. Quiescence and rebellion in an Appalachian valley,* University of Illinois Press.

Ge X., Xu C., Misaki D., Markus H. R., Tsai J. L. (2024, May) How culture shapes what people want from AI. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*, Honolulu, HI, USA. ACM.

Ghoshal D. (2017, March 28) The world's largest biometric ID programme is a privacy nightmare waiting to happen. *Quartz India.* Retrieved from https://qz.com/943102/aadhaar-for-dummies-why-right-thinking-indians-should-be-worried-over-the-slow-death-of-privacy/

Ghoshal S. (2017, August 3) The bitter reality of being a transgender employee in India. *HuffPost.* Retrieved from https://www.huffingtonpost.in/2017/07/19/the-bitter-reality-of-being-a-transgender-employee-in-india_a_23035751/

Goel V. (2018, April 7) 'Big Brother' in India requires fingerprint scans for food, phones, and finances. *New York Times.* Retrieved from https://www.nytimes.com/2018/04/07/technology/india-id-aadhaar.html

Greenleaf G. (2010) India's national ID system: Danger grows in a privacy vacuum. *Computer Law & Security Report,* vol. 26, no 5, pp. 479–491.

Hameed S. (2020, January 13) Government, UIDAI helpless as instances of Aadhaar misuse reach terrifying new levels. *The Noose.* Retrieved from https://timesofindia.indiatimes.com/readersblog/the-noose/government-uidai-helpless-as-instances-of-aadhaar-misuse-reach-terrifying-new-levels-9686/

Hayward C. R. (1998) De-facing power. *Polity,* vol. 31, no 1, pp. 1-22.

Henne K. (2019) Surveillance in the name of governance: Aadhaar as a fix for leaking systems in India. *Information, technology and control in a changing world* (B. Haggart, K. Henne, N. Tusikov eds.), Palgrave Macmillan, pp. 197-215.

HK V. (2018, September 26) SC's Aadhaar verdict: Privacy vs identity. *Deccan Herald.* Retrieved from https://www.deccanherald.com/india/aadhaar-act-verdict-history-693614.html

Hoofnagle C. J., van der Sloot B., Zuiderveen Borgesius F. (2019) The European Union general data protection regulation: What it is and what it means. *Information & Communications Technology Law,* vol. 28, no 1, pp. 65-98.

Human Rights Watch (2018, January 13) India: Identification project threatens rights. Retrieved from https://www.hrw.org/news/2018/01/13/india-identification-project-threatens-rights

Islam N. (2019) Edward W. Said's *Orientalism*: A thematic interpretation. *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 6, no 6, p. 159.

Jabalaji S. (2017, June 7) Indian technology companies support Aadhaar through secure services. *Forbes*. https://www.forbes.com/sites/sindhujabalaji/2017/06/07/indian-technology-companies-support-aadhaar-through-secure-services/

Johnson R. B., Onwuegbuzie A. J. (2004) Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, vol. 33, no 7, pp. 14–26.

Kaushik, M. (2010). Configuring the UID. *Business Today*, 19(22), 12.

Khaira R. (2018, January 3) Rs 500, 10 minutes, and you have access to billion Aadhaar details. *The Tribune*. Retrieved from http://www.tribuneindia.com/news/nation/rs-500-10-minutes-and-you-have-access-to-billion-aadhaar-details/523361.html

Khera R. (2019) *Dissent on Aadhaar: Big data meets Big Brother,* Orient BlackSwan.

Kim S.-W., Lee Y. (2023) Investigation into the influence of socio-cultural factors on attitudes toward artificial intelligence. *Education and Information Technologies.* Advance online publication. https://doi.org/10.1007/s10639-023-12172-y

Kulesz O. (2017) Cultural policies in the age of platforms. *Re|shaping cultural policies: Advancing creativity for development* United Nations Educational, Scientific and Cultural Organization, pp. 73–88. https://unesdoc.unesco.org/ark:/48223/pf0000260592

Kulesz O. (2022) Culture, platforms, and machines: The impact of artificial intelligence on the diversity of cultural expressions. *Artificial Intelligence and Culture: Perspectives for Cultural Diversity in the Digital Age* (Brazilian Internet Steering Committee — CGI.br, pp. 39–66. Retrieved from https://cetic.br/media/docs/publicacoes/7/20221111151258/sectoral_studies-artificial_intelligence_and_culture.pdf

Kulkarni S. (2023, October 20) This new Aadhaar-related banking fraud is on the rise; why you need to lock your Aadhaar biometrics now. *The Economic Times*. https://economictimes.indiatimes.com/tech/technology/aadhar-data-leak-personal-data-of-81-5-crore-indians-on-sale-on-dark-web-report/articleshow/104856898.cms

Kumar A. (2023) Indian technology and culture in the context of globalization. *RP World: History and Culture Studies,* no 2, pp. 21–27.

Kumar H. S., Madan S. (2020) A study on Aadhaar privacy and personal security issues in India. *Purakala*, vol. 31, no 11, p. 72.

Li T. M. (2007) Governmentality. *Anthropologica*, vol. 49, no 2, pp. 275–281. Canadian Anthropology Society.

Loomba A. (1998) *Colonialism/postcolonialism,* Routledge.

Loomba A. (2014) *Colonialism/postcolonialism* [PDF file], Routledge. http://armytage.net/pdsdata/%5BAnia_Loomba%5D_Colonialism_Postcolonialism_(The_New(Book4You).pdf

LotusArise (2024, May 2) Theories of power. *LotusArise*. https://lotusarise.com/theories-of-power/

McKinsey (2010) Inclusive growth and financial security. Retrieved from http://ccmrm.org/wp-content/uploads/2015/05/McKinsey-2010-inclusive-growth-report.pdf

McLeod J. (2000) *Beginning postcolonialism,* Manchester University Press.

McPhee-Knowles S., Kanar L. (2022) Review of *Decolonizing data: Unsettling conversations about social research methods* by J. M. Quinless, University of Toronto Press, p. 151.

Means A. J. (2022) Foucault, biopolitics, and the critique of state reason. *Educational Philosophy and Theory,* vol. 54, no 12.

Meloni M. (2023) An unproblematized truth: Foucault, biopolitics, and the making of a sociological canon. *Social Theory & Health,* vol. 21, no 1, pp. 99–118.

Ministry of Electronics and Information Technology (2023) *Digital personal data protection act 2023.* https://www.meity.gov.in/writereaddata/files/Digital%20Personal%20Data%20Protection%20Act%202023.pdf

Mira U. B., Kar A. K., Dwivedi Y. K., Gupta M. P., Sharma R. S. (2019) Realizing digital identity in government: Prioritizing design and implementation objectives for Aadhaar in India. *Government Information Quarterly,* vol. 36, no 4, p. 101442.

Moneylife Digital Team (2023, May 2) Proliferation of fake Aadhaar cards a threat to integrity of 2024 elections: EAS Sarma. *Moneylife.* https://www.moneylife.in/article/proliferation-of-fake-aadhaar-cards-a-threat-to-integrity-of-2024-elections-eas-sarma/69688.html

Moosavinia S. R., Niazi N., Ghaforian A. (2011) Edward Said's *Orientalism* and the study of the self and the other in Orwell's *Burmese Days*. *Studies in Literature and Language,* vol. 2, no 1, pp. 103–113.

Mukherjee A., Nayar L. (2011, December 5) Aadhar — A few basic issues. *Outlook India.* Retrieved from http://www.outlookindia.com/article/Aadhar-A-Few-Basic-Issues/279077

Murphy L., Maguire W. (2011) Applying mixed methods research in evaluating clinical trials. *Qualitative Research in Accounting & Management,* vol. 8, no 1, pp. 72–90.

Nair A., Eskici B. (2023) Digital public services: The development of biometric authentication in India. *Introduction to development engineering: A framework with applications from the field* (T. Madon, A. J. Gadgil, R. Anderson, L. Casaburi, K. Lee, A. Rezaee eds.), Springer International Publishing, pp. 533–561.

Narayanan P., Ramanathan M. (2023) Assessing Aadhaar: India's unique identification project in the light of global experiences. *Journal of South Asian Development,* vol. 18, no 1, pp. 20–35.

O'Callahan T. (2020, March 27) What happens when a billion identities are digitized? *Yale Insights.* Retrieved from https://insights.som.yale.edu/insights/what-happens-when-billion-identities-are-digitized

Okolo C. T. (2023) *AI explainability in the Global South: Towards an inclusive praxis for emerging technology users* (Doctoral dissertation, Cornell University). https://doi.org/10.13140/RG.2.2.25596.51841

Okolo C. T., Dell N., Vashistha A. (2022, June) Making AI explainable in the Global South: A systematic review. *Proceedings of the 2022 ACM Conference on Computing and Sustainable Societies (COMPASS '22),* Seattle, WA, USA. Association for Computing Machinery. https://doi.org/10.1145/3530190.3534802

Pandey A. (2023, August 19) Aadhaar and the rise of digital financial fraud in India. *LiveMint.* Retrieved from https://www.livemint.com/news/india/aadhaar-and-the-rise-of-digital-financial-fraud-in-india-11692673835964.html

Pandya A. (2020) Aadhaar and India's biometric experiment: Implications for privacy and data security. *Journal of Asian Studies,* vol. 79, no 4, pp. 957–984.

Pat R. K., Kumar V., Jain N. (2015, February 13–15) Positives, negatives, and future of AADHAAR: A project management perspective. *PMI India Research & Academic Conference 2015,* Mumbai, India.

Powell C. S. (2016) Surveillance normalization. *Harvard Civil Rights-Civil Liberties Law Review,* vol. 58.

PowerCube (2023) Power and the politics of surveillance: The PowerCube framework.

Press Information Bureau (2021, November 16) Pratyaksh Hanstantrit Labh (PAHAL) scheme — An overview. Retrieved from https://static.pib.gov.in/WriteReadData/specificdocs/documents/2021/nov/doc2021111621.pdf

PRS India (2023) *The Digital Personal Data Protection Bill, 2023.* Ministry of Electronics and Information Technology. Retrieved from https://prsindia.org/billtrack/digital-personal-data-protection-bill-2023

Quinless J. M. (2022) *Decolonizing data: Unsettling conversations about social research methods,* University of Toronto Press.

Rabinow P. (1991) *The Foucault reader,* Pantheon Books.

Rabinow P. (Ed.) (1991) *The Foucault reader: An introduction to Foucault's thought,* Penguin.

Ra S., Pandey A. (2023, August 24) Unseen and unrecognised: The Indians excluded from Aadhaar. *Haqdarshak.* Retrieved from https://haqdarshak.com/2023/08/24/unseen-and-unrecognised-the-indians-excluded-from-aadhaar/

Raju R. S., Singh S., Khatter K. (2017) Aadhaar card: Challenges and impact on digital transformation. *Department of Computer Science and Engineering, Manav Rachna International University, Faridabad, Haryana-121004, India; Accendere Knowledge Management Services Pvt. Ltd., Chennai-600101, India.* https://doi.org/10.48550/arXiv.1708.05117

Rama A. (2020, January 17) The Aadhaar card: Cybersecurity issues with India's biometric experiment. *Jackson School of International Studies, University of Washington.* https://jsis.washington.edu/news/the-aadhaar-card-cybersecurity-issues-with-indias-biometric-experiment/

Ranjan A., Kumar N. (2024) The politics of Aadhaar: State surveillance, digital identities, and marginalization in India. *South Asian Journal of Political Science,* vol. 42, no 2, pp. 110–126.

Ranjan P. (2015) Edward Said's Orientalism: A post-colonial culture study. *IOSR Journal of Humanities and Social Science (IOSR-JHSS,* vol. 20, no 9, pp. 85–88.

Sadhya D., Sahu T. (2024) A critical survey of the security and privacy aspects of the Aadhaar framework. *Computers & Security,* no 140, p. 103782.

Said E. W. (1978) *Orientalism,* Pantheon Books.

Said E. W. (1979) *Orientalism,* Vintage Books.

Said E. W. (1993) *Culture and imperialism,* Chatto & Windus.

Said E. W. (2001) *Orientalism*, Penguin Books.

Sarma E. A. S. (2023, April 29) Proliferation of fake Aadhaar cards a threat to integrity of 2024 elections in India. *Countercurrents.* Retrieved from https://countercurrents.org/2023/04/proliferation-of-fake-aadhaar-cards-a-threat-to-integrity-of-2024-elections/

SC Observer. (2017, August 24) Judgment of the Court in Plain English (I): Fundamental right to privacy. *SC Observer.* Retrieved from https://www.scobserver.in/reports/k-s-puttaswamy-right-to-privacy-judgment-of-the-court-in-plain-english-i/

Sethi A. (2018, March 13) Why state data hubs pose a risk to Aadhaar security. *Hindustan Times.* Retrieved from https://www.hindustantimes.com/india-news/why-state-data-hubs-pose-a-risk-to-aadhaar-security/story-Klyl3yT5MkFk6Szg2yGg9N.html

Sharma P. (2018, March 15) Only male or female can get PAN card, transgender told. *Times of India.* Retrieved from https://timesofindia.indiatimes.com/india/only-male-or-female-can-get-pan-card-transgender-told/articleshow/63321785.cms

Sharma S., Hill M. (2024, April 26) The biggest data breach fines, penalties, and settlements so far. *CSO Online.* Retrieved from https://www.csoonline.com/article/567531/the-biggest-data-breach-fines-penalties-and-settlements-so-far.html

Shiva H. K., Madan, S. (2021) A study on Aadhaar-based digital identity: Technological advancements and privacy concerns. *Information Technology for Development,* vol. 27, no 3, pp. 455–470.

Singh A. (2024, July 20) From removing Muslim staff to displaying Aadhaar cards, how dhaba owners are complying with Kanwar Yatra order. *CNN-News18.* https://www.cnnnews18.com/india/from-removing-muslim-staff-to-displaying-aadhaar-cards

Singh M. (2017, February 14) Is India's central database with biometric details of its billion citizens a privacy nightmare? *Mashable.* Retrieved from https://mashable.com/2017/02/14/india-aadhaar-uidai-privacy-security-debate/#ZgQoT5iIwOqc

Smith L. T. (1999) *Decolonizing methodologies: Research and indigenous peoples,* Zed Books.

Sobti S., Sahni B., Bala K. (2020) Surgical coverage of cataract in a rural area of north India: A cross-sectional study. *Journal of Family Medicine and Primary Care,* no 9, p. 4112.

Soja E. W. (1996) *Thirdspace: Journeys to Los Angeles and other real-and-imagined places,* Blackwell.

Spivak G. C. (1988) Can the subaltern speak? *Marxism and the interpretation of culture* (C. Nelson, L. Grossberg eds.), Macmillan Education, pp. 271–313.

Strange S. (1998) *States and markets,* Pinter.

Sundquist M. (2023) *The Age of Surveillance: Global Trends in Biometric* Identification Systems, Cambridge University Press.

Supreme Court Observer. (2018, September 26) Constitutionality of Aadhaar Act: Judgment summary. Retrieved from https://www.scobserver.in.

Supreme Court of India. (2021, January 11) *Justice K. S. Puttaswamy (Retd.) & Anr. vs. Union of India & Ors.* Retrieved from https://main.sci.gov.in/supremecourt/2018/45777/45777_2018_5_1001_25344_Judgement_11-Jan-2021.pdf

Tashakkori A., Teddlie C. (eds.). (2003) *Mixed methods in social and behavioral research,* Sage Publications Inc.

Thaker A. (2018, May 1) Aadhaar's mixing of public risk and private profit. *The Caravan.* Retrieved from https://caravanmagazine.in/reportage/aadhaar-mixing-public-risk-private-profit

The Times of India (2024, January 10) Aadhaar and the challenges of biometric authentication in India. *The Times of India.* https://timesofindia.indiatimes.com/india/aadhaar-and-the-challenges-of-biometric-authentication-in-india/articleshow/104312567.cms

Totapally S., Sonderegger P., Rao P., Gosselt J., Gupta G. (2019) *State of Aadhaar: A people's perspective,* New York: Dalberg.

Tseng A. S., Edwards C., Rawls M., McGinn M., Wieghard N., Santen S. A., Deiorio N. M. (2023) A mixed methods exploration of the emergence of goal orientation in medical students' individualized learning plans. *Medical Teacher,* no 45, pp. 588–595.

Unique Identification Authority of India (n.d.) Vision and mission. Retrieved August 10, 2024, from https://uidai.gov.in/en/about-uidai/unique-identification-authority-of-india/vision-mission.html

Venkatanarayanan A. (2017, December 6) The 360 degree database. *Kaarana.* Retrieved from https://medium.com/karana/the-360-degree-database-17a0f91e6a33

Venkatanarayanan A. (2018) Aadhaar enrolment costs. Retrieved July 4, 2019, from https://medium.com/karana/aadhaar-enrolment-costs-bc17f0d30018

Aadhaar, искусственный интеллект и идентичность: взаимодействие власти и надзора на Глобальном Юге

Ниланджана Синха
Доцент Бизнес-школы NSHM
Адрес: БЛ Саха Роуд, 124, Калькутта — 700 053 Западная Бенгалия, Индия
E-mail: nilanjana.sinha@gmail.com

Первоначально задуманный как инструмент управленческого совершенствования и содействия социальной инклюзии, Aadhaar[1] стал сочетанием инструментов, несущим серьезные последствия для индийского общества. Включенный в исследование основанной

1. Двенадцатизначный личный идентификационный номер для жителей Индии на основе биометрических и демографических данных

на технологиях системы государственного администрирования, Aadhaar раскрывает запутанный механизм, который ставит соответствующие вопросы, касающиеся надзора, управления и идентичности. В исследовании, опирающемся на интервью, анализ обработки данных и постколониализм, ставится цель показать работу Aadhaar в индийском обществе, разобрав его парадоксальное действие одновременно в качестве инструмента управления и идентификационного надзора.

Несмотря на обещание Aadhaar сделать льготы легкодоступными, его практические возможности представляют собой проблему, например, из-за наличия дубликатов удостоверений личности или активных учетных записей умерших людей. Подобное развитие событий вызывает соответствующие опасения по поводу массовой слежки и приватизации государственных функций частными компаниями, такими как OnGrid и Khosla Labs, что, в свою очередь, ставит важные вопросы этики и подотчетности. По сравнению с другими глобальными режимами, такими как GDPR или Social Security Systev, в Aadhaar отсутствуют строгие механизмы защиты частной жизни физических лиц и обеспечения подотчетности и всех необходимых уровней прозрачности.

Надзорные механизмы Aadhaar, основанные на концепции паноптикона у Фуко, нормализуют возможности государства по мониторингу и контролю граждан при помощи биометрической слежки. Это переводит инклюзию в область новой формы управления идентичностью, где те, кто обладает меньшими шансами на одобрение и признание — например, мигранты, бедняки, лица, лишенные доступа к банковским услугам — снова оказываются в аналитически более управляемой и удобной среде социального исключения. Понятие гибридности из работ Х. Бабы позволяет по-иному взглянуть на действие Aadhaar в постколониальном контексте: оно изменчиво и должно постоянно переопределяться в процессе надзора. Критический постколониальный анализ, проведенный Эдвардом Саидом, показывает, как Aadhaar мимикрирует под колониальные системы категоризации, еще больше усиливая властные иерархии и систематическое, институциональное исключение.

По мере того, как Aadhaar все больше формирует индийскую идентичность, он делает это не просто как инструмент социальной инклюзии, а как система, которая систематически помещает граждан под надзор, превращая их идентичность в синоним проверки и регулирования, нормализуя власть слежки и контроля без всякого их выбора либо согласия.

Ключевые слова: Aadhaar, биометрическая идентификация, надзор, идентичность, постколониальная теория

doi: 10.17323/1728-192x-2024-4-113-139

Digital Orientalism in Machine Vision: A Cross-Platform Analysis of AI-Generated Representations of Indian Culture

Jinu K. Varghese
Research Fellow at Manipal Institute of Communication, Manipal Academy of Higher Education(MAHE), Manipal, Karnataka, India
Address: Madhav Nagar, Manipal — 576104 Karnataka, India
E-mail: jinukvarghese7@gmail.com

Padma Rani
Professor & Director of Manipal Institute of Communication, Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, India
Address: Madhav Nagar, Manipal — 576104 Karnataka, India
E-mail: padma.rani@manipal.edu

This study investigates how contemporary artificial intelligence image generation systems interpret and reproduce Indian cultural elements through a comparative analysis of three major platforms: Stable Diffusion, Flux, and Midjourney. While these systems have demonstrated remarkable technical capabilities, their handling of non-Western cultural elements remains understudied. We present a novel methodological framework that combines visual social semiotics with digital anthropology to analyse AI-generated images across multiple dimensions, including representational accuracy, cultural sensitivity, and power dynamics. Our systematic analysis of images generated through increasingly sophisticated prompts reveals significant patterns in how these systems mediate cultural representation. Results indicate that while platforms exhibit varying technical proficiencies, they consistently demonstrate concerning biases in human representation, particularly in their treatment of gender, class, and ethnic identities. The analysis reveals systematic simplification of complex cultural elements and the persistence of orientalist perspectives, despite advances in technical capabilities. These findings suggest that improved technical sophistication alone is insufficient for authentic cultural reproduction; rather, fundamental reconsideration of how these systems process and understand cultural information is necessary. This research contributes both theoretical insights into digital cultural representation and practical implications for developing more culturally sensitive AI systems, while highlighting crucial areas for improvement in the technical architecture of image generation models.

Keywords: Digital Orientalism, Artificial Intelligence, Cultural Representation, Visual Social Semiotics, Indian Culture, Machine Vision, Digital Anthropology, Postcolonial Computing

Introduction

The rapid evolution of artificial intelligence image generation systems has introduced unprecedented capabilities in visual content creation, while simultaneously raising critical questions about how these systems interpret and reproduce cultural elements. As AI-generated imagery becomes increasingly prevalent in global visual culture, understanding how these systems represent non-Western cultures becomes crucial for both technical development and cultural preservation. This study examines how three prominent AI image generation platforms; Stable Diffusion, Flux, and Midjourney (Ticong,

2024; Pandit, 2024) interpret and reproduce Indian cultural elements, revealing complex patterns of what might be termed "digital orientalism."

The emergence of AI image generation systems marks a significant shift in how cultural representations are produced and circulated globally. These systems, trained on vast datasets of human-created images, have become powerful mediators of cultural representation, capable of generating complex visual interpretations of cultural elements with minimal human input (Qadri et al., 2023; Baum & Villasenor, 2024). However, their role in cultural representation, particularly in the context of interpreting non-Western cultures, remains largely unexplored, with only a few studies addressing this aspect. This gap becomes particularly significant when considering India, a culture whose visual traditions often exist in tension with Western modes of representation. Contemporary AI image generation systems operate at the intersection of technical capability and cultural interpretation, raising fundamental questions about how computational systems understand and reproduce cultural elements. While these systems demonstrate remarkable technical sophistication in generating visual content, their handling of cultural nuance, particularly regarding non-Western cultures, remains problematic (Qadri et al., 2023; Ghosh et al., 2024).

Building upon existing scholarship in digital cultural representation, this research extends analysis into the domain of generative AI. Previous studies have examined how digital technologies mediate cultural representation (Hall et al., 2024; Taylor, 1996), but the emergence of AI image generation systems introduces new complexities requiring systematic investigation. Through analysis of visual patterns, power dynamics, and cultural authenticity across varying levels of prompt complexity, this study reveals how technical capabilities influence representation quality. Through systematic analysis of 270 AI-generated images, this study employs a novel methodological framework combining visual social semiotics with digital anthropology to examine how these systems handle cultural complexity. The research reveals that while technical capability influences representation quality, achieving authentic cultural representation may require fundamental reconsideration of how these systems understand and process cultural information.

Review of Literature

The emergence of AI image generation systems has introduced new complexities in cultural representation, necessitating a critical examination of how these technologies interpret and reproduce cultural elements. Existing scholarship provides crucial frameworks for understanding these dynamics while revealing significant gaps in our understanding of AI-mediated cultural representation.

Cultural Representation in Digital Spaces

Edward Said's (1978) foundational work on Orientalism established critical frameworks for understanding how Western perspectives shape representations of non-Western cul-

tures. While Said's analysis focused on traditional media and academic discourse, his insights into how power structures influence cultural representation take on new significance in the age of artificial intelligence. Scholars have begun applying Said's frameworks to digital spaces (Roh et al., 2015; Nakamura, 2002), yet limited attention has been paid to how AI image generation systems specifically perpetuate or transform orientalist perspectives.

Nakamura's (2002) work on "cybertypes" introduced crucial concepts for understanding how digital technologies reproduce and reinforce racial and cultural stereotypes. Her research demonstrates how seemingly neutral technological systems can encode and perpetuate existing cultural biases. However, her work predates the emergence of sophisticated AI image generation systems, leaving open questions about how these new technologies might transform or amplify the patterns she identified. Recent scholarship has attempted to bridge this gap. Benjamin's (2019) concept of the "New Jim Code" and Noble's (2018) work on algorithmic oppression provide valuable frameworks for understanding how AI systems perpetuate cultural biases. Yet these analyses primarily focus on search engines and classification systems rather than generative AI technologies.

Recent empirical work has begun to specifically examine how AI image generation systems interpret and reproduce cultural elements. Of particular significance, Qadri et al.'s (2023) community-cantered study of text-to-image models in South Asia predominantly employs qualitative focus group methodology, leaving quantitative and systematic cross-platform analysis largely unexplored. Their broad examination of South Asian representation, while valuable, doesn't address the specific technical mechanisms through which different AI platforms interpret and reproduce cultural elements. The absence of comparative platform analysis leaves unanswered questions about how varying technical capabilities might influence cultural representation quality. While Qadri et al. (2023) 's study identifies problematic patterns in cultural representation, it doesn't systematically analyse how these patterns might vary across different levels of prompt complexity or technical sophistication. The relationship between technical capability and representation quality remains unclear, particularly regarding how increased complexity might affect cultural authenticity. A significant gap also exists in understanding intra-cultural dynamics within specific regional contexts. While broad examinations of South Asian representation provide valuable insights, detailed analysis of how AI systems handle internal cultural complexities — particularly regarding caste, class, and regional diversity within specific national contexts — remains limited. The lack of categorical analysis across different aspects of cultural representation leaves our understanding of AI's handling of cultural nuance incomplete.

Building upon Qadri et al.'s work, Ghosh et al. (2024) conducted focus groups with diverse Indian subcultural participants to examine text-to-image generators' impact on cultural representation, documenting novel harms of exoticism and cultural misappropriation. However, their study was limited by its focus on qualitative data without cross-platform comparisons, lack of structured visual analysis methodology, and insufficient investigation of how technical elements and prompt complexity influence cultural repre-

sentation accuracy. These limitations highlight the need for more comprehensive technical analysis alongside cultural sensitivity in studying AI-generated representations.

Technical Foundations and Cultural Implications

The technical evolution of AI image generation systems has been accompanied by growing awareness of their cultural implications. Crawford and Paglen's (2021) research on training datasets reveals how historical biases become embedded in AI systems' foundational materials. Their examination of ImageNet demonstrates how classification systems inherently reproduce cultural hierarchies yet doesn't specifically address how these biases manifest in generative systems. Shankar et al.'s (2017) work on geographic sampling bias in dataset collection provides crucial insights into why AI systems struggle with non-Western contexts. Their research reveals systematic errors in recognition systems when dealing with underrepresented geographic and cultural regions. However, their focus on recognition rather than generation leaves open questions about how these biases influence creative AI systems.

The relationship between training data and cultural interpretation has emerged as a critical area of concern. Prabhu and Birhane's (2020) analysis of large image datasets reveals troubling patterns in how standard training data perpetuates problematic representations of marginalized communities. While their work provides valuable insights into dataset bias, it doesn't fully address how these biases transform when filtered through generative AI systems.

Gender and Intersectional Perspectives

The examination of gender, intersectionality in AI-generated imagery reveals significant gaps in current research frameworks. While Butler's (2002) theory of gender performativity and Klein's and D'Ignazio (2024) concept of "data feminism" provide valuable theoretical foundations, their application to AI-generated visual representations remains limited. Current research has not fully explored how AI systems specifically encode and reproduce gender norms through visual elements, particularly in non-Western contexts. Although Hill Collins' (2022) "matrix of domination" framework helps understand intersecting oppressions, its application to analysing AI-generated imagery's reproduction of multiple, simultaneous inequalities requires further development.

Mohanty's (1988) critique of Western feminist discourse's homogenization of "third world women" becomes increasingly relevant, yet current research hasn't adequately examined how this homogenization manifests in AI systems' visual interpretations. While studies acknowledge AI systems' Western-centric training data, detailed analysis of how this affects gender representation across different cultural contexts remains unexplored. The systematic reproduction of gender stereotypes across AI platforms lacks comprehensive comparative analysis, particularly regarding how different technical capabilities influence gender representation in cultural contexts.

Visual Analysis and Digital Culture

The examination of cultural representation through visual analysis has gained new complexity with the emergence of AI-generated imagery. The application of visual social semiotics, pioneered by Kress and van Leeuwen's (2006) seminal work "Reading Images: The Grammar of Visual Design," provides fundamental frameworks for understanding how meaning emerges through visual elements. Their analysis of vectors, modality, and compositional structures remains crucial for understanding how AI systems interpret and reproduce cultural signifiers. However, their framework, developed for traditional visual media, requires significant adaptation to address the unique characteristics of algorithmically generated imagery.

Recent scholars have begun applying visual social semiotics to digital cultural representation. Jewitt and Oyama's (2004) work on "Visual Meaning: A Social Semiotic Approach" demonstrates how digital technologies transform traditional meaning-making processes. Their analysis of how power relations manifest in digital imagery provides valuable insights, though their work primarily focuses on human-created digital content rather than AI-generated imagery. Similarly, O'Halloran's (2013) research on "Multimodal Analysis and Digital Technology" explores how digital platforms influence visual meaning-making but doesn't specifically address the role of artificial intelligence in cultural representation.

The complexity of cultural representation in digital spaces finds important theoretical grounding in Rose's (2016) "Visual Methodologies." Her comprehensive framework for analysing site of production, image itself, and site of reception becomes particularly relevant when examining AI-generated content. However, her methodology requires expansion to address how AI systems simultaneously function as both producers and interpreters of visual culture. Mirzoeff's (2015) concept of "visual culture 2.0" provides crucial context for understanding digital visual production yet doesn't fully account for the algorithmic mediation of cultural elements.

Scholars examining cultural representation in digital spaces have highlighted the importance of contextual analysis. Pauwels' (2021) work on "A Multimodal Framework for Analysing Websites as Cultural Expressions " provides valuable frameworks for analysing digital cultural artifacts, though his methodology doesn't specifically address AI-generated content. Zhao and Zappavigna's (2018) research demonstrate how social semiotics can be adapted for digital content analysis, but their work primarily focuses on social media and selfie analysis rather than AI-generated imagery.

The intersection of technology and visual culture finds important theoretical development in Mackenzie and Munster's (2019) analysis of machine vision systems. Their work reveals crucial insights into how computational systems process visual information, though they don't specifically examine how different AI architectures might influence cultural interpretation. This gap becomes particularly relevant when considering how varying technical capabilities might affect cultural representation across different AI platforms.

Colour semiotics in digital spaces, examined extensively in Van Leeuwen's (2011) work, provides crucial insights into how cultural meaning manifests through colour

choices. His analysis becomes particularly relevant when examining how AI systems interpret and reproduce culturally specific colour associations, though his framework requires adaptation for analysing algorithmic colour interpretation. This connects with Aiello's (2020) work on visual semiotics, which provides valuable frameworks for understanding cultural meaning-making in digital spaces, though her concept of "visual citizenship" needs expansion to fully address AI-generated imagery.

More recent scholarship has begun examining multimodal analysis in digital contexts. Bateman, Wildfeuer, and Hiippala's (2017) comprehensive framework for analysing multimodal digital content provides valuable tools for understanding how different semiotic modes interact. However, their work doesn't fully address how AI systems integrate and interpret multiple cultural modes simultaneously. The role of platform governance in shaping visual cultural production, examined in Gillespie's (2018) work, provides important context for understanding how technical decisions influence cultural representation. His analysis helps explain how platform architecture shapes cultural interpretation, though his framework requires expansion to address the specific challenges of AI-generated imagery. This connects with Dourish's (2022) examination of digital materiality, which offers valuable perspectives on how digital systems process cultural information, though his work doesn't specifically address the role of AI architectures in this process.

These scholars collectively demonstrate the rich potential of visual social semiotics for analysing digital cultural representation, while also revealing significant gaps in our understanding of how these frameworks apply to AI-generated imagery. The need for adapted methodological frameworks that can address the unique characteristics of algorithmic cultural interpretation becomes increasingly apparent, particularly regarding how technical capabilities influence visual meaning-making processes.

The existing literature highlights significant gaps in understanding how AI systems interpret and reproduce cultural representations. Foundational works, such as Said's Orientalism (1978) and Nakamura's analyses of digital stereotypes (2002), offer critical frameworks for studying cultural representation. However, their applicability to emerging AI technologies, particularly text-to-image generative systems, remains insufficiently explored. While recent scholarship has begun to address issues of bias and representational patterns in AI, these studies predominantly focus on classification systems rather than the distinct challenges posed by generative AI technologies.

Text-to-image AI systems introduce unique complexities in cultural representation that challenge existing theoretical models. For instance, Qadri et al.'s (2023) investigation into South Asian representation in AI systems illuminates the phenomenon of "algorithmic orientalism," yet it leaves open questions regarding how different technical architectures shape cultural interpretation. Similarly, the interplay between technical sophistication and cultural authenticity, specifically how varying levels of prompt complexity influence representational quality, remains understudied.

Although scholarship has extensively examined bias in AI systems and media portrayals of Indian culture, limited attention has been devoted to the mediation of cultural representation by AI image generators. Existing analytical frameworks for visual

culture and digital bias require adaptation to account for the distinct characteristics of AI-generated imagery. Furthermore, the intersection of technical capabilities, cultural interpretation, and embedded power dynamics in AI-generated visual representations remains inadequately theorized. Crucially, the extent to which technical sophistication either mitigates or exacerbates problematic representations remains unexplored, particularly when addressing cultural elements that necessitate nuanced understanding of historical, social, and religious contexts.

This study seeks to address these critical gaps by conducting a systematic analysis of how three prominent AI platforms interpret and reproduce Indian cultural elements. By examining visual patterns, power dynamics, and cultural authenticity across varying levels of prompt complexity, the research extends existing theoretical frameworks into the domain of AI-generated imagery. This approach not only offers new insights into the relationship between technical capability and cultural interpretation but also provides practical recommendations for developing culturally sensitive AI systems.

Methodology

This research employs a comprehensive mixed-methods approach to examine AI-generated representations of Indian culture across Midjourney (Version 6.1), Flux (FLUX 1.1), and Stable Diffusion (3.5 Large). Our methodological framework combines systematic data collection with rigorous visual social semiotic analysis, enabling detailed examination of how these platforms interpret and reproduce cultural elements.

Data Collection and Sampling

Data collection was conducted between September-November 2024, using standardized settings across all platforms to ensure consistency. The study implements a structured data collection protocol across five primary categories (Geographic Locations, Cultural Practices/Traditions, Socioeconomic Conditions, Prominent Figures/Leaders, and Diversity). This categorical approach is justified by the need to comprehensively capture different aspects of Indian cultural representation. Within each category, we developed 15 prompts at three distinct complexity levels (simple, medium, and detailed), generating a total dataset of 270 images (90 images per platform). Simple prompts contained basic descriptive elements, medium prompts incorporated specific cultural and contextual details, while detailed prompts included complex cultural nuances and intersectional elements. Prompts used are given in the annexture. This graduated approach to prompt complexity enables examination of how technical sophistication influences cultural representation quality.

This graduated approach generated a total dataset of 270 images (90 per platform), with standardized image specifications (1024x1024 resolution, PNG/JPEG format) across all platforms. Midjourney generates four images by default for each prompt. To maintain consistency and ensure a streamlined analysis process, we selected the first image from each output for this study. This approach eliminates potential bias

in choosing between images, provides a consistent basis for comparison across all prompts, and simplifies the methodology while retaining the representational qualities of the platform's output.

The study's classification of traditional versus modern elements was based on several key indicators. Traditional elements were defined as cultural markers with historical continuity of over 50 years, including classical art forms, religious symbols, traditional dress (e.g., saris, dhotis), architectural styles (e.g., temple architecture), and traditional social structures. Modern elements encompassed contemporary cultural expressions, including fusion fashion, modern architecture, technological integration, and evolving social dynamics. This categorization framework enables systematic analysis of how AI systems negotiate between historical and contemporary representations of Indian culture.

Analytical Framework

The analytical framework employed in this study builds upon and adapts traditional visual social semiotics to address the unique challenges of analysing AI-generated cultural representations. This adaptation is necessitated by the complex intersection of technical capabilities, cultural interpretation, and power dynamics in AI-generated imagery.

Three-Tier Analytical Framework

The study implements a novel three-tier analytical framework that enables comprehensive examination of both technical and cultural elements.

The first tier, Visual Pattern Analysis, builds upon Kress and van Leeuwen's visual grammar to examine compositional elements, colour usage, spatial arrangements, and technical aspects. The framework has been specifically modified for AI-generated content through the addition of algorithmic artifact analysis, integration of platform-specific technical markers, and adaptation of compositional analysis for machine-generated imagery.

The second tier, Power Dynamics Analysis, draws from Butler's gender performativity theory and Hill Collins' matrix of domination to examine gender representation and roles, class hierarchies and socioeconomic markers, cultural power structures, and intersectional dynamics across social categories.

The third tier, Cultural Representation Analysis, informed by Said's Orientalism and Mohanty's postcolonial critique, evaluates cultural authenticity markers, stereotyping patterns, identity construction, and religious symbolism.

Cross-Platform Comparative Analysis

The study's methodological innovation lies in its systematic cross-platform comparison approach:

- Analysis of identical prompts across platforms enables examination of how technical capabilities influence cultural interpretation.

- Varying prompt complexity levels (simple, medium, detailed) reveals how technical sophistication affects representation quality.
- Categorical examination across five domains (Geographic Locations, Cultural Practices/Traditions, Socioeconomic Conditions, Prominent Figures/Leaders, and Diversity) provides structured comparison points.

This methodological framework addresses limitations in existing approaches by enabling simultaneous analysis of technical and cultural aspects, responding to Mackenzie and Munster's call for integrated studies of machine vision systems. Its systematic cross-platform comparison extends beyond single-platform analyses to reveal how technical architectures shape cultural representation and how technical capabilities influence representation quality. By analysing complexity levels, the framework addresses gaps in understanding how technical sophistication affects cultural representation, offering insights into the interplay between capability and quality. Its comprehensive cultural analysis moves beyond identifying issues to uncovering their technical and cultural origins, examining how AI systems navigate cultural complexity.

The framework's adaptability and replicability make it a valuable methodological tool. Its structured approach supports analysis of diverse cultural contexts, technical and cultural dimensions and provides protocols for comparative studies. By adapting visual social semiotics for AI-generated content, the framework advances methodological rigor, enabling deeper understanding of how AI systems interpret and reproduce cultural elements across architectures. It establishes a robust foundation for exploring the relationship between technical capability and cultural representation in AI-generated imagery.

Coding and Analysis Process

Images underwent rigorous coding using a comprehensive scheme encompassing visual elements, cultural markers, power indicators, stereotyping patterns, and technical quality metrics. Two independent coders analysed each image to ensure reliability, with disagreements resolved through discussion with a third coder. (The coding scheme, Prompts, Categories and Complexity are being given in the annexture)

Analysis and Discussion

Visual Pattern Analysis

Analysing 270 AI-generated images using visual social semiotics reveals complex patterns in how AI systems interpret and reproduce Indian cultural elements. These patterns, when examined through Kress and van Leeuwen's (2006) visual grammar framework, demonstrate how AI systems perpetuate what might be termed "algorithmic orientalism", the systematic reproduction of Western perspectives and power dynamics through computational systems.

The prevalence of centered compositions in AI-generated imagery (Stable Diffusion 45%, Flux 42%, Midjourney 40%) exemplifies what Kress and van Leeuwen identify as "visual sa-

lience," while the limited 20% representation of hierarchical layouts demonstrates what Jewitt and Oyama (2001) term "compositional meaning." This dual pattern reveals how AI systems not only reproduce Western compositional paradigms but also fail to fully capture traditional Indian visual hierarchies. While Mitter (2001) emphasizes Indian art's use of distributed narratives and hierarchical scaling, the standardized outputs of AI platforms suggest a systematic bias toward Western ways of seeing (Berger, 1972), indicating deeper limitations in these systems' ability to interpret and reproduce culturally diverse visual traditions.

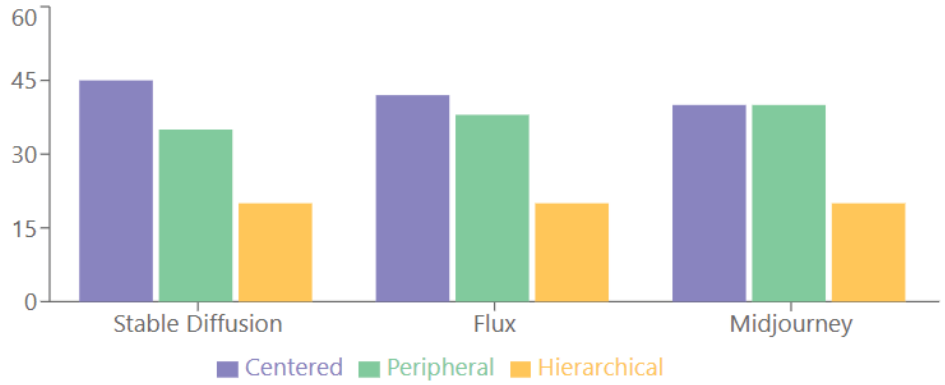


Figure 1. Compositional Patterns Across Platforms

The systematic disparity between modern and traditional colour representations in AI platforms (40% vs 30% in both Flux and Midjourney) reveals a significant pattern in how these systems approach cultural authentication. Traditional Indian colour palettes, as Doshi (2023) notes, are deeply rooted in cultural symbolism and regional diversity, with colours carrying specific ritual and social significance that extends beyond mere aesthetic value. These traditional colours, which Doshi (2023) identifies as integral to India's cultural identity through their use in festivals, rituals, and artistic practices, are consistently underrepresented compared to modern interpretations that Favor saturated, digital-friendly hues. Through Kress and van Leeuwen's framework of visual modality, this colour treatment indicates not just aesthetic preference but claims about visual truth and authenticity. As van Leeuwen (2011) argues, colour choices carry deep cultural and social meanings beyond mere aesthetics, making this modernization of traditional Indian colour schemes particularly significant. O'Halloran's (2013) concept of "semiotic dissonance" helps explain how this 10% gap between modern and traditional colour palettes represents a systemic disconnect in how AI platforms interpret and reproduce cultural visual elements. The preference for contemporary colour schemes over traditional Indian pigmentation demonstrates how AI systems may inadvertently dilute cultural authenticity in pursuit of modern aesthetic standards. Furthermore, Aiello's (2020) framework of "visual citizenship"

suggests these colour disparities reflect broader patterns of cultural inclusion and exclusion in AI-generated imagery, where the higher representation of modern palettes (40%) potentially marginalizes traditional visual languages and their associated cultural meanings.

The distribution of gaze patterns across AI platforms (38% direct, 42% indirect, 20% no gaze) reveals significant insights about algorithmic representation and cultural power dynamics. Through Kress and van Leeuwen's framework of "interactive meanings," the predominance of indirect gaze (42%) points to a systematic pattern in how AI systems mediate viewer-subject relationships. This aligns with Nakamura's (2002) analysis of "cybertypes," where digital representations perpetuate colonial patterns of looking, the higher percentage of indirect gaze suggests an unconscious reproduction of traditional orientalist visualization practices. Rose's (2016) concept of "scopic regimes" helps explain how this gaze distribution reflects culturally specific ways of seeing that have been embedded in AI systems. Furthermore, Mirzoeff's (2015) framework of "visual subalternity" illuminates how the significant presence of indirect gaze (42%) and the limited representation of direct engagement (38%) potentially reinforces the marginalization of non-Western visual traditions in digital spaces, suggesting a deeper structural bias in how AI systems interpret and reproduce cultural ways of seeing.

The conspicuous absence of digital devices and infrastructure in AI-generated market and youth scenes (75% showing no devices) reveals a problematic pattern in algorithmic representation. Contemporary urban Indian spaces, characterized by ubiquitous digital payment systems, ATMs, digital signage, and personal devices, are consistently rendered by AI systems in a technologically sanitized manner. This erasure exemplifies what Mackenzie and Munster (2019) term "algorithmic visuality," where AI systems reproduce an ahistorical vision of cultural spaces. The minimal representation of modern technology (15% personal devices, 10% professional/public) and complete absence of digital infrastructure (ATMs, digital signboards, QR codes) in market scenes aligns with Dourish's (2022) concept of "digital materiality," revealing how AI systems embed outdated cultural assumptions in their foundational architectures. This systematic technological erasure, particularly in scenes where digital devices and infrastructure would be commonplace in contemporary Indian life, reflects what Benjamin (2019) terms the "New Jim Code," where algorithmic systems perpetuate stereotypical, pre-digital representations of non-Western spaces through seemingly neutral design choices, effectively divorcing cultural subjects from their contemporary technological realities.

The analysis of directional elements in AI-generated images reveals significant limitations in how these systems interpret and reproduce traditional Indian visual narratives. Through Bateman et al.'s (2017) framework of "multimodal coherence," the preference for static compositions over dynamic ones is particularly evident in group images, where characters uniformly face the viewer, eliminating the internal vectors and dynamic relationships that typically characterize interpersonal interactions. As Dehejia (2010) notes in "Indian Art," traditional pictorial narratives employ complex directional elements and interpersonal gazes to convey multilayered stories and relationships, yet AI systems consistently default to frontal, viewer-directed compositions that eliminate these internal dynamics. This standardization of gaze and the notable absence of inter-character vectors flattens the compositional complexity and reduces the rich relational storytelling to more simplified, presentational formats.

Figure 2. Platforms used Flux (Image 1), Stable Diffusion (Image 2) and Midjourney (Image 3) (in the order of appearance) Prompt ("A group of Indian men engaging in household chores and caregiving tasks")

The distribution of dress representations in AI-generated images (45% traditional, 30% modern, 15% hybrid) reveals significant patterns in how algorithms encode and reproduce cultural and gender identities. Through Pauwels' (2012) framework of "cultural scripts," the dominance of traditional dress forms suggests a standardized, potentially reductive approach to representing Indian identity. This becomes particularly evident in the system's problematic handling of gender roles, where AI automatically adds feminine cultural markers (saris, bangles) when depicting men in domestic roles. Instead of representing domestic labour as universal human activity, the system has internalized societal biases that fundamentally link household work with feminine presentation, exemplifying what Zhao and Zappavigna (2018) term "visual stereotyping." The significant gap between traditional (45%) and hybrid dress representations (15%) reflects Klein's and D'Ignazio (2024) concept of "data colonialism," demonstrating how AI systems perpetuate colonial perspectives and gender hierarchies simultaneously. These patterns reveal how AI training data and architectures entrench both cultural

and gender binaries, making it nearly impossible for the system to conceptualize either contemporary Indian dress practices or domestic work in nuanced, non-stereotypical ways.

These findings extend beyond simple digital orientalism to reveal what might be termed "algorithmic cultural reduction", the systematic simplification of complex cultural visual grammars through computational processes. The consistency of patterns across platforms suggests what Gillespie (2018) identifies as "platform governance", how technical architectures themselves embed cultural assumptions and biases.

Power Dynamics Analysis

When generating images of Indian social interactions, AI systems reveal deeply embedded assumptions about who holds power and how it is displayed. Each platform consistently encodes specific biases about gender roles, class markers, and professional status, though they do so in notably different ways.

The distribution of gender representation across platforms reveals what Kress and van Leeuwen (2006) term "representational structures" that perpetuate traditional power hierarchies. Stable Diffusion's seemingly balanced distribution (35% male, 30% female, 35% non-binary/no human element) masks deeper structural inequalities when analysed through Butler's (2002) framework of gender performativity. The predominance of male figures in positions of authority and modern professional settings, juxtaposed against female representation in traditional and domestic contexts, suggests what Mohanty (1988) identifies as the perpetuation of colonial gender hierarchies through technological means. The progression across platforms, from Stable Diffusion's 35% male representation to Midjourney's 40% indicates what Klein's and D'Ignazio (2024) term "power asymmetries" in algorithmic systems. The progression to 40% male representation in Midjourney suggests a concerning trend where AI systems are amplifying existing gender biases in Indian society. What appears particularly telling is how these systems consistently place men in positions of modern authority (like business leaders or tech professionals) while relegating women to traditional roles (like domestic settings or cultural ceremonies), essentially digitizing the traditional Indian gender divide through modern technology.

The representation of social class across AI platforms reveals complex patterns in what Mirzoeff (2011) terms "scopic regimes" — historically specific ways of seeing and representing that shape how power and visibility operate in visual culture. While Stable Diffusion initially appears balanced with equal representation of middle and working classes (30% each), this superficial equity masks deeper biases when examined through Said's (1978) orientalist framework. Three key patterns emerge: First, the consistent correlation of elite markers (25%) with modern, urban settings perpetuates what Benjamin (2019) terms "coded bias" — the systematic reproduction of class hierarchies through technical systems. Second, Midjourney's progression toward privileged representation (35% middle class, 30% elite) demonstrates what Gillespie (2018) identifies as "platform values," showing how technical architectures embed social hierarchies. Third, and most concerning, the reduction in working-class representation to 20% reflects what Prabhu and Birhane (2020) term "representational harm," effectively

erasing marginalized groups through algorithmic systems. This distortion becomes particularly problematic in the Indian context, where the majority of the population belongs to working-class backgrounds yet finds limited representation in AI-generated imagery.

Figure 3. Flux (image 1), Stable Diffusion (image 2) and Midjourney (image 3) (in the order of appearance)- Prompt ("A group portrait highlighting the achievements and contributions of Indian scientists, engineers, and technologists from diverse backgrounds")

The analysis of professional activities reveals how AI systems interpret India's traditional-modern occupational divide, though these findings are limited to our specific set of prompts. The dominance of traditional occupations in Stable Diffusion (35%) demonstrates what Jewitt and Oyama (2001) identify as "modal affordances" — the ways

technical systems interpret and reproduce cultural signifiers. While Flux shows an equal distribution between traditional and modern occupations (30% each), this apparent balance masks deeper hierarchies when examined through Mackenzie and Munster's (2019) framework of algorithmic visuality. Notably, Stable Diffusion's representation of traditional craftspeople, street vendors, and agricultural workers often romanticizes these roles while diminishing their contemporary relevance. Similarly, Flux's seemingly balanced distribution fails to capture how traditional occupations actively evolve within modern Indian society, presenting them instead as static, unchanged practices.

The intersection of professional activities with gender and class markers in AI-generated images reveals deeply embedded patterns of social bias. Through Mirzoeff's (2015) concept of "visual subalternity" and Mohanty's (1988) critique of Western feminist discourse, we see how these systems perpetuate both colonial and patriarchal views of Indian professional identity. The consistent dominance of male figures in modern professional settings, coupled with the disproportionate representation of women in traditional roles, reflects Hill Collins' (2000) "matrix of domination" in digital spaces. This gendered professional divide demonstrates not only Crawford and Paglen's (2021) "classification politics" but also exemplifies Nakamura's (2002) concept of "cybertypes" — how digital systems crystallize and perpetuate cultural stereotypes. The AI's systematic linking of modernity with masculinity and tradition with femininity illustrates Benjamin's (2019) observations about how seemingly neutral technical systems encode social hierarchies. These representations become particularly problematic in contemporary India, where professional gender roles are actively being challenged and redefined, yet AI systems continue to reinforce outdated power structures through what Klein's and D'Ignazio (2024) identify as "technological redlining" of gender and professional identity.

These findings suggest that AI systems, despite their technological sophistication, are not merely reproducing but actively amplifying existing social hierarchies within Indian society. The consistent patterns across platforms reveal fundamental challenges in representing contemporary Indian social structures, raising critical questions about how these technologies shape cultural perceptions both locally and globally.

Cultural Representation Analysis

Our analysis reveals how AI platforms consistently reduce India's complex cultural fabric to oversimplified, often exoticized imagery. Drawing from postcolonial theory, we find these systems not only reproduce but actively reshape cultural narratives through a distinctly Western lens.

The representation of religious imagery in AI-generated content reveals systematic biases in how these systems interpret Indian spiritual diversity. The dominance of Hindu symbols (45% across platforms, reaching 48% in Stable Diffusion) exemplifies Said's (1978) orientalist framework, where India's complex religious fabric is reduced to its most globally recognizable elements. This creates what might be termed a "digital Hindu aesthetic," marginalizing other faiths through limited representation (15% Islamic, 10% other traditions) — a pattern reflecting Spivak's (1988) concept of "epistemic violence" through digital erasure.

Moreover, the significant absence of religious elements (30%) in everyday contexts demonstrates what Miller and Horst (2020) call "cultural flattening," failing to capture how religion permeates daily Indian life. This bias perpetuates through what Appadurai (1996) terms "cultural circulation," creating a feedback loop where Western perceptions shape training data, which in turn reinforces oversimplified representations of Indian religious identity.

The overwhelming presence of traditional symbols (45%) compared to modern (25%) and hybrid (15%) elements demonstrates what Bhabha (1994) terms "temporal fixing", freezing colonial subjects in an imagined traditional past. In AI-generated representations of Indian culture, this manifests as a preference for showcasing traditional elements (like classical dance forms or traditional dress) over contemporary cultural expressions, essentially creating a "digital museum" version of Indian culture rather than capturing its living, evolving nature.

Figure 4. Platforms used Flux (image 1), Stable Diffusion (image 2) and Midjourney (image 3) (in the order of appearance)- Prompt ("A scene of an interfaith religious ceremony, where priests and spiritual leaders from Hinduism, Islam, Sikhism, Christian and other faiths come together to perform rituals")

Figure 5. Flux (image 1), Stable Diffusion (image 2), Midjourney (image 3) and original Kathakali performance (image 4) — creative common licence (in the order of appearance)- Prompt ("A troupe of Kathakali performers in full costume and elaborate makeup, enacting a scene from the Mahabharata")

The representation of cricket in AI-generated imagery provides a compelling example of temporal displacement in algorithmic representation. Despite India's status as a global cricket powerhouse with modern facilities and thriving cricket culture, AI systems consistently generate anachronistic imagery anchored in a colonial aesthetic — from outdated clothing to antiquated playing fields. This misrepresentation is compounded by geographic bias in AI development, where systems trained primarily on U.S.-centric datasets fail to accurately capture the sport's contemporary reality in South Asia and other Commonwealth nations. The resulting imagery reflects what might be termed "algorithmic colonialism," where AI systems not only exoticize modern cricket-playing nations

but also systematically misrepresent the sport's technical sophistication. This dual distortion exemplifies Bhabha's (1994) concept of colonial stereotype, now manifesting in algorithmic space through what appears as neutral technology but actually perpetuates and amplifies colonial-era visual tropes.

Figure 6. Flux (image 1), Stable Diffusion (image 2) and Midjourney (image 3) (in the order of appearance) — Prompt ("Indian men playing a game of cricket, watched by a crowd of enthusiastic spectators")

The relationship between prompt complexity and cultural representation reveals systemic biases in AI image generation. As Benjamin (2019) terms "default discrimination," simpler prompts trigger a high prevalence of traditional imagery (75%), showing how these systems default to orientalist stereotypes without explicit guidance. This bias manifests differently across platforms, exemplifying what Nakamura (2002) identifies as "cybertypes" — digital spaces' tendency to reproduce racial and cultural stereotypes.

For instance, while Stable Diffusion accurately renders architectural details, it struggles with religious diversity, suggesting AI systems can simultaneously preserve and distort cultural elements. The notably limited representation of hybrid cultural elements (15%) demonstrates what Horst and Miller (2021) term "digital cultural lag," where AI systems fail to capture the dynamic interplay between tradition and modernity in contemporary Indian culture. This technological inability to represent cultural fluidity reinforces artificial boundaries between traditional and modern elements that rarely exist in lived experience.

Figure 7. Flux (image 1), Stable Diffusion (image 2) and Midjourney (image 3) (in the order of appearance)- Prompt ("An Indian railway station bustling with activity, where wealthy passengers board a high-speed train while low-income workers unload goods from carts nearby")

While this study focuses on broad patterns of Indian cultural representation, we acknowledge the limitations in capturing India's vast regional diversity. The analysis reveals how AI systems often default to pan-Indian cultural markers, potentially overlooking regional nuances. This is particularly evident in the representation of dress (45% traditional, 30% modern, 15% hybrid), where regional variations in traditional attire are often simplified into nationally recognizable forms. These patterns reveal that AI systems' struggle with cultural complexity stems not merely from technical limitations but from fundamental assumptions encoded within their architecture. The consistent defaulting to traditional imagery and inability to represent hybrid cultural elements shows how these systems actively reinforce rigid cultural categories rather than capturing the fluid, evolving nature of contemporary Indian culture. This suggests a critical need to reconsider how AI systems are trained to interpret and represent cultural information.

Cross Platform Analysis

The comparative analysis across Stable Diffusion, Flux, and Midjourney reveals distinct patterns in how different technical architectures interpret and reproduce Indian cultural elements. Each platform demonstrates unique strengths and limitations that illuminate broader challenges in AI-based cultural representation.

Stable Diffusion exhibits higher adherence to traditional colour schemas (45% compared to Flux's 40% and Midjourney's 35%), demonstrating what O'Halloran (2013) terms "semiotic fidelity" in digital spaces. While this fidelity suggests technical sophistication, it inadvertently reinforces what Said (1978) identifies as orientalist perspectives through over-emphasis of conventional cultural markers. In contrast, Flux shows superior capability in integrating technological elements (32% compared to Stable Diffusion's 25% and Midjourney's 28%), better capturing the hybrid reality of modern Indian life where traditional and technological elements coexist. Midjourney's sophisticated handling of spatial relationships (40% balanced compositions) reflects Kress and van Leeuwen's (2006) concept of "compositional meaning," though often prioritizing aesthetic balance over cultural authenticity.

The relationship between prompt complexity and output quality reveals significant patterns across platforms. Simple prompts consistently trigger high percentages of traditional dress representation (Stable Diffusion 75%, Flux 72%, Midjourney 70%), demonstrating systematic bias in default outputs. However, with medium-complexity prompts, Flux shows notable improvement with a 35% reduction in stereotypical elements, compared to Stable Diffusion's 30% and Midjourney's 28%, suggesting greater algorithmic adaptability to nuanced cultural contexts.

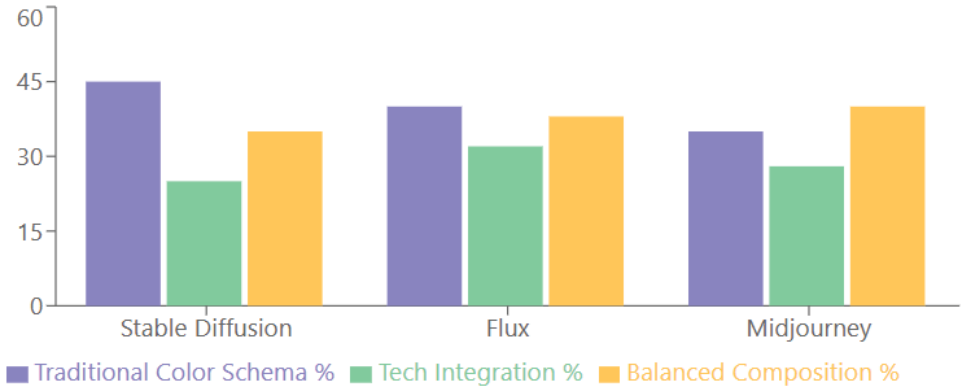


Figure 8. Technical Capabilities Across Platforms

Figure 9. Platform-Midjourney, Prompt-left side image ("A lively Garba dance performance during the Navratri festival, with dancers in traditional ghaghra choli costumes"), Prompt-right side image ("An Indian wedding where the bride and groom come from different religious and regional backgrounds")

Perhaps most concerning is Midjourney's consistent tendency to "whitewash" Indian subjects, depicting them with lighter skin tones and European features — a bias less prevalent in Stable Diffusion and Flux. This pattern exemplifies what Qadri et al. (2023) identify as "prompt-dependent cultural literacy," where systems require explicit guidance to avoid defaulting to Western-centric representations. While Midjourney demonstrates superior handling of social interactions (80% accuracy), its struggle with professional hierarchies (65% traditional occupations in working-class representations) indicates that technical sophistication in one area doesn't necessarily translate to comprehensive cultural understanding.

The platforms' handling of intersectional elements further reveals systemic limitations in representing complex cultural identities. Stable Diffusion's consistent 30% female representation across social classes demonstrates what might be termed "performative uniformity" — a formulaic approach to gender representation that fails to capture the nuanced gender dynamics within different Indian social contexts. While Flux shows more sophisticated handling of class intersectionality, maintaining 30% working-class representation in urban settings, its struggle with religious diversity (only 8% non-Hindu symbols) indicates persistent limitations in representing India's multi-religious fabric.

These intersectional variations suggest what we might call "algorithmic cultural variance," where different technical approaches lead to distinctly different interpretations of Indian culture. While increased prompt complexity generally improves representation quality, as Qadri et al. (2023) note through their concept of "prompt-dependent cultural literacy," it also reveals the systems' fundamental reliance on explicit guidance to move beyond orientalist defaults. Even Midjourney's seemingly sophisticated handling of social interactions (80% accuracy) is undermined by its struggle with professional hierarchies, particularly in representing working-class occupations (65% traditional representations).

These patterns collectively demonstrate that while technical capabilities vary across platforms, all three systems share fundamental limitations in processing cultural complexity. Each platform demonstrates distinct technical approaches to cultural representation. Stable Diffusion shows stronger architectural accuracy but struggles with religious diversity. Flux demonstrates superior handling of hybrid cultural elements but shows limitations in representing intersectional identities. Midjourney excels in compositional balance but tends to westernize physical features. These variations suggest that technical capabilities influence cultural representation in complex and sometimes contradictory ways. This suggests a critical need for "cultural translation" in AI architectures — developing systems that can authentically represent the nuanced interplay of gender, class, and religion in Indian society without defaulting to reductive stereotypes. Such development becomes increasingly crucial as these platforms' interpretations increasingly shape global perceptions of cultural narratives.

Conclusion

This study's examination of AI-generated representations of Indian culture reveals how technological advancement alone does not guarantee authentic cultural representation. Through analysis of Stable Diffusion, Flux, and Midjourney, we find that these systems do not merely reflect but actively reshape cultural narratives through computational processes. The consistent patterns of digital exoticization, intersectional blindness, and prompt-dependent representation suggest fundamental limitations in how AI systems process cultural information.

Our findings demonstrate that while increased technical sophistication and prompt complexity can yield more nuanced cultural representations, the underlying challenge

lies not in technical capabilities but in how these systems fundamentally conceptualize culture. The tendency to default to orientalist perspectives, particularly in representing gender, class, and religious intersections, points to deeper issues in AI architecture and training data that cannot be resolved through technical refinement alone.

Limitations

This study's examination of AI-generated Indian cultural representation faced several methodological and technical constraints. The analysis was limited to three major AI platforms, potentially missing patterns present in other systems. The sample size of 270 images, while substantial for qualitative analysis, may not capture the full range of possible representations. While this study focuses on broad patterns of Indian cultural representation, we acknowledge the limitations in capturing India's vast regional diversity. The research was also constrained by the temporal limitation of current AI systems and their training data, which may not reflect very recent cultural developments. Additionally, the analysis relied on predetermined categorical frameworks for coding visual elements, which might have missed nuanced cultural meanings that fall outside these categories.

Future Directions

Future research in AI-generated cultural representation should explore several critical dimensions to address current limitations. The field would benefit from expanded analysis of emerging AI platforms, particularly those developing in non-Western contexts, as well as deeper investigation into how these systems handle regional variations within Indian culture, especially less represented traditions.

Annexture

https://docs.google.com/spreadsheets/d/1Oyb4PaFrHXDYpCjXjo3zYFc_CsBawps4Sl1d-lFo1tUY/edit?usp=sharing

References

Aiello G. (2020) Visual semiotics: key concepts and new directions. *The SAGE Handbook of Visual Research Methods* (Second Edition), London: SAGE Publications, pp. 367-380.

Appadurai A. (1996) *Modernity at Large: Cultural Dimensions of Globalization*, Minneapolis: University of Minnesota Press.

Barthes R. (1977) *Image-Music-Text*, London: Macmillan.

Bateman J. A., Wildfeuer J., Hiippala T. (2017) *Multimodality: Foundations, Research and Analysis — A Problem-Oriented Introduction*, Berlin: De Gruyter Mouton.

Baum J., Villasenor J. (2024) Rendering Misrepresentation: Diversity Failures in AI Image Generation. *Brookings*. Available at: www.brookings.edu/articles/rendering-misrepresentation-diversity-failures-in-ai-image-generation (accessed 30 November 2024).

Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610–623.

Benjamin R. (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*, Medford: Polity Press.

Bhabha H. K. (1994) *The Location of Culture* (2nd ed.), London: Routledge.

Bourdieu P. (2018) Distinction: A Social Critique of the Judgement of Taste. *Food and Culture*, pp. 287-318.

Butler J. (2006) *Gender Trouble: Feminism and the Subversion of Identity* (1st ed.), London: Routledge.

Chakrabarty D. (2000) *Provincializing Europe: Postcolonial Thought and Historical Difference*, Princeton: Princeton University Press.

Chatterjee P. (1993) *The Nation and Its Fragments: Colonial and Postcolonial Histories*, Princeton: Princeton University Press.

Crawford K., Paglen T. (2021) Excavating AI: The Politics of Images in Machine Learning Training Sets. *AI & Society*. https://doi.org/10.1007/s00146-021-01162-8

Dehejia V. (1997) *Indian Art*, London: Phaidon.

Doshi S. (2022) The Influence of Culture, Evolving Symbolisms and Globalization on Defining Colour Forecasting in India. *Fashion, Style & Popular Culture*, vol. 9, no 1, pp. 9-26.

Dourish P. (2022) *The Stuff of Bits: An Essay on the Materialities of Information*, Cambridge: MIT Press.

Dulhanty C., Wong A. (2019) Auditing ImageNet: Towards a Model-Driven Framework for Annotating Demographic Attributes of Large-Scale Image Datasets. arXiv. https://doi.org/10.48550/ARXIV.1905.01347

Eubanks V. (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York: St. Martin's Press.

Ghosh S., Venkit P., Gautam S., Wilson S. (2024) Do Generative AI Models Output Harm while Representing Non-Western Cultures: Evidence from A Community-Centered Approach. *7th AAAI/ACM Conference on AI, Ethics, and Society*. Available at: https://arxiv.org/pdf/2407.14779v2. (accessed 30 November 2024).

Gill R. (2016) Post-postfeminism?: New Feminist Visibilities in Postfeminist Times. *Feminist Media Studies*, vol. 16, no 4, pp. 610–630.

Gillespie T. (2019) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*, New Haven: Yale University Press.

Gray M. L., Suri S. (2019) *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*, Boston: Eamon Dolan Books.

Hall S., Nixon S., Evans J. (eds.) (2024) *Representation: Cultural Representations and Signifying Practices*, London: SAGE Publications.

Hill Collins P. (2000) Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment (2nd ed.), London: Routledge.

Horst H., Miller D. (eds.) (2012) *Digital Anthropology*, London: Routledge.

Irani L., Vertesi J., Dourish P., Philip K., Grinter R. E. (2010) Postcolonial Computing: A Lens on Design and Development. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1311–1320. https://doi.org/10.1145/1753326.1753522

Jewitt C., Oyama R. (2004) Visual Meaning: A Social Semiotic Approach. *The Handbook of Visual Analysis*, London: SAGE Publications, pp. 134-156.

Klein L., D'Ignazio C. (2024) Data Feminism for AI. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, New York: Association for Computing Machinery, pp. 100-112. Available at: https://doi.org/10.1145/3630106.3658543

Kress G., Leeuwen T. V. (2020) *Reading Images: The Grammar of Visual Design* (3rd ed.), London: Routledge.

MacKenzie A., Munster A. (2019) Platform Seeing: Image Ensembles and Their Invisualities. *Theory, Culture & Society*, vol. 36, no 5, pp. 3-22.

Mankekar P. (2015) *Unsettling India: Affect, Temporality, Transnationality*, Durham: Duke University Press.

Mirzoeff N. (2011) The Right to Look. *Critical Inquiry*, vol. 37, no 3, pp. 473–496.

Mirzoeff N. (2015) *How to See the World*, London: Pelican Books.

Mitchell M. (2019) *Artificial Intelligence: A Guide for Thinking Humans*, New York: Farrar, Straus and Giroux.

Mohamed S., Png M.-T., Isaac W. (2020) Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. https://doi.org/10.48550/ARXIV.2007.04068

Mohanty C. T. (1988) Under Western Eyes: Feminist Scholarship and Colonial Discourses. *Feminist Review*, no 30, pp. 61–88.

Muldoon J., Wu B. A. (2023) Artificial Intelligence in the Colonial Matrix of Power. *Philosophy & Technology*, vol. 36, no 80.

Nakamura L. (2002) *Cybertypes: Race, Ethnicity, and Identity on the Internet* (1st ed.), London: Routledge.

Noble S. U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York: NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5

O'Halloran K. (2013) Multimodal Analysis and Digital Technology. *Readings in Intersemiosis and Multimedia* (ed. E. Montagna), Israel: IBIS Editions, pp. 35-53.

Pandit B. (2024) Flux AI Image Generator: A Guide with Examples. *DataCamp*. Available at: www.datacamp.com/tutorial/flux-ai (accessed 30 November 2024).

Pauwels L. (2012) A Multimodal Framework for Analyzing Websites as Cultural Expressions. *Journal of Computer-Mediated Communication*, vol. 17, no 3, pp. 247-265.

Prabhu V. U., Birhane A. (2020) Large Image Datasets: A Pyrrhic Win for Computer Vision?. *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1536-1546. https://doi.org/10.48550/ARXIV.2006.16923

Qadri R., Shelby R., Bennett C., Denton E. (2023) AI's Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia. *Proceedings of*

the 2023 ACM Conference on Fairness, Accountability, and Transparency, New York: Association for Computing Machinery, pp. 506-517.

Rani G., Singh J., Khanna A. (2023) Comparative Analysis of Generative AI Models. *2023 International Conference on Advances in Computation, Communication and Information Technology*, pp. 760–765. https://doi.org/10.1109/ICAICCIT60255.2023.10465941

Roh D. S., Huang B., Niu G. A. (eds.) (2015) *Techno-Orientalism: Imagining Asia in Speculative Fiction, History, and Media*, New Brunswick: Rutgers University Press.

Rose G. (2016) *Visual Methodologies: An Introduction to Researching with Visual Materials*, London: SAGE Publications.

Said E. W. (1978) *Orientalism*, New York: Vintage Books.

Shankar S., Halpern Y., Breck E., Atwood J., Wilson J., Sculley D. (2017) No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World. arXiv. Available at: https://arxiv.org/abs/1711.08536

Spivak G. (1988) Can the Subaltern Speak?. *Marxism and the Interpretation of Culture* (eds. C. Nelson, L. Grossberg), Urbana: University of Illinois Press, pp. 271-313.

Tacheva J., Ramasubramanian S. (2023) AI Empire: Unraveling the Interlocking Systems of Oppression in Generative AI's Global Order. *Big Data & Society*. https://doi.org/10.1177/20539517231219241

Taylor J. (1996) New Media and Cultural Representation. *Information Society. Human-centred Systems* (ed. K. Gill), London: Springer, pp. 265-282. https://doi.org/10.1007/978-1-4471-3249-3_16

Ticong L. (2024) Midjourney vs Stable Diffusion: 2024's Creative Clash. *eWEEK*. Available at: www.eweek.com/artificial-intelligence/midjourney-vs-stable-diffusion (accessed 30 November 2024).

Van Leeuwen T. (2011) *The Language of Colour: An Introduction*, London: Routledge.

Zhao D., Wang A., Russakovsky O. (2021) Understanding and Evaluating Racial Biases in Image Captioning. arXiv. https://doi.org/10.48550/ARXIV.2106.08503

Цифровой ориентализм в машинном зрении: кроссплатформенный анализ репрезентаций индийской культуры, сгенерированных искусственным интеллектом

Джину К. Варгезе
Научный сотрудник Манипальского института коммуникаций
Манипальской академии высшего образования,
Адрес: Мадхав Нагар, Манипал — 576104 Карнатака, Индия
E-mail: jinukvarghese7@gmail.com

Падма Рани
Профессор и директор Манипальского института коммуникаций
Манипальской академии высшего образования,
Адрес: Мадхав Нагар, Манипал — 576104 Карнатака, Индия
E-mail: padma.rani@manipal.edu

Данное исследование рассматривает интерпретацию и воспроизводство элементов индийской культуры современными системами формирования изображений, основанными на искусственном интеллекте, на примере сравнительного анализа трех крупных ИИ-платформ: Stable Diffusion, Flux и Midjourney. Несмотря на примечательные технические возможности, продемонстрированные данными системами, их обращение с элементами незападных культур остается недостаточно изученным. Мы представляем новую методологическую рамку из визуальной социальной семиотики и цифровой антропологии для анализа сгенерированных искусственным интеллектом изображений по множеству параметров, таких как точность репрезентации, учет культурных особенностей и динамика отношений власти. Проведенный нами систематический анализ изображений, генерируемых в ходе постоянно усложняющихся запросов, выявляет наличие значительных шаблонов, возникающих при репрезентации культуры данными системами. Результаты исследования говорят о том, что, несмотря на различающееся техническое совершенство, показываемое этими платформами, они всякий раз демонстрируют предвзятость в производстве человеческих образов, особенно в части их гендерной, классовой и этнической идентичности. Анализ раскрывает систематическое упрощение сложных культурных элементов и неизменно сохраняющуюся ориенталистскую перспективу, несмотря на развитие технических возможностей. Предлагаемые выводы свидетельствуют о том, что для настоящего воспроизводства культуры недостаточно одного лишь технического совершенства; необходимо скорее фундаментальное переосмысление того, как данные системы обрабатывают и понимают культурную информацию. Данное исследование предлагает как теоретический вклад в вопросы цифровой репрезентации культур, так и практические выводы для разработки более внимательных к культурным особенностям систем искусственного интеллекта, а также показывает важные области для совершенствования технической архитектуры моделей генерации изображений.

Ключевые слова: цифровой ориентализм, искусственный интеллект, репрезентация культуры, визуальная социальная семиотика, индийская культура, машинное зрение, цифровая антропология, постколониальное вычисление

doi: 10.17323/1728-192x-2024-4-140-172

The Impact of Socio-Cultural and Demographic Factors on Gen AI Accessibility, Usability, and Applicability in the UAE

Aizhan Shomotova
College of Interdisciplinary Studies, Zayed University
Address: P. O. Box 144534, Abu Dhabi, UAE
E-mail: aizhan.shomotova@zu.ac.ae

Salwa Husain
College of Interdisciplinary Studies, Zayed University
Address: P. O. Box 144534, Abu Dhabi, UAE
E-mail: salwa.husain@zu.ac.ae

Areej Elsayary
College of Interdisciplinary Studies,
Address: P. O. Box 144534, Abu Dhabi, UAE
E-mail: areej.elsayary@zu.ac.ae

Arindam Das
Professor, School of Liberal Arts
Alliance University
Address: Chikkahagade Cross, Chandapura-Anekal Main Road
Anekal, Bangalore — 562106, Karnataka, India
E-mail: arindam.das@alliance.edu.in

Glenn Muschert
Professor, Department of Public Health and Epidemiology
College of Medicine and Health Sciences, Khalifa University or Science and Technology
Address: P.O. Box 127788, Abu Dhabi, UAE
E-mail: glenn.muschert@ku.ac.ae

Despite rapid AI integration in education globally, the role of socio-cultural and demographic factors in shaping AI adoption in diverse settings like the UAE remains underexplored. This study explores the accessibility, usability, and applicability of generative AI (Gen AI) tools, such as ChatGPT, in higher education, focusing on Emirati university students. Employing a mixed-methods research design, data were collected from 83 students through demographic surveys and structured reflections based on Gibbs' Reflective Cycle (Gibbs, 1988). Quantitative analysis showed strong links between how Gen AI uses AI and demographic and educational factors, such as high school type, gender, and academic performance, with a slight trend seen for the father's education level. Students from private international schools reported higher daily usage, while females showed greater weekly usage frequencies, emphasizing the influence of educational background and gender on technology adoption. Thematic analysis identified Gen AI accessibility as a key theme, with students highlighting the instant feedback and time-saving benefits of Gen AI tools. However, disparities in digital literacy and infrastructural challenges remain. The Gen AI Usability Study showed that students liked the detailed feedback and how it broke down complex ideas. However, they ran into problems like prompt sensitivity and the need to verify output. The Gen AI Applicability results emphasized ChatGPT's help in creating content, improving language, and organizing academic material, but they also pointed out that it has trouble following assignment-specific rules. This study contributes to the literature by addressing socio-demographic factors influencing Gen AI adoption in a culturally diverse setting like the UAE. The results make it clear

that Gen AI literacy and critical evaluation skills must be taught in schools, and strategies are necessary to ensure everyone has equal access (Ben Romdhane et al., 2023). Recommendations aim to help educators leverage Gen AI's potential while mitigating its challenges in higher education.

Keywords: generative AI, socio-demographics, education, UAE, usage, accessability

Introduction

The spread of generative artificial intelligence (Gen AI) technologies worldwide has led to much discussion about using them in different social and cultural settings, especially in higher education. The United Arab Emirates (UAE) has emerged as a pioneering force in AI adoption, demonstrating substantial progress and investment in AI technologies, mainly through its ambitious UAE AI Strategy 2031 (Shamout, Ali, 2021). It is the socio-cultural and demographic background of the students, that shape the students' perspective, had been considered one of the fundamentally important bases for a successful implementation and utilization of AI system in the educational context (Chan, Ho, 2023).

From the perspective of the Global South, the appropriateness of the study is underscored by the unique socio-cultural and economic context that goes to give a basis to and create a structure for AI adoption in the UAE and the Gulf region. The multicultural context of the Global South and the highly structured levels of infrastructure with reference to technology conditions a unique paradigm for integrating generative AI (Gen-AI) in higher education. Our study that primarily focuses on students of the university level in UAE projects the latters' usability, usefulness, and accessibility specially when conditioned through socio-cultural and demographic factors. This comparatively micro-understanding projects our focused epistemy around AI at a macro level to comprehend the Global South situation in general. The straggling of traditional cultural values and technological progress that so defines the value system of UAE is a highly acceptable understanding generally for any Global South situation. The results of our study indicate the need for culturally-sensitive-AI-integration-strategies that takes into cognizance the peculiar geo-political spaces, intrinsic opportunities, and impending challenges.

With UAE's recent foray into the world of technology that is meant to transform its educational system, the necessity to comprehend a student's lived experience and perceptions becomes imperative. More recent research points towards a UAE undergrad's socio-cultural demographic backgrounds and subjective experiences that go to impact the undergrad's navigation of AI-based technologies (Yusuf et al., 2024). A combination of qualitative and quantitative approach in the previously mentioned research has enabled a probing into the dynamics of the UAE undergrad students' responses to open-ended questions in this study and further providing comprehensive understanding with an undergrad's engagement and interaction with AI-enabled technologies in their academic trajectory.

The higher education system of UAE and its incorporation of AI had strategically made an approach to make a foray unto an improvement into the learning-based outcome and experiences of its undergrad students (Johnson et al., 2022). Yet for, Bamatraf et al. (2021) the understanding of the socio-economic implication of AI from a students' vantage point emphasize its cruciality for effective implementation and adaption in educational settings. Here one may be directed towards an interesting guiding framework that facilitates a reflective analysis of students' perspectives — Gibbs' Reflective Cycle (Gibbs, 1988). The Gibbs-ian model helps to structure and interpret, in a qualitative manner, a student's engagement with Gen-AI across 06 stages *description, feelings, imagination, analysis, conclusion,* and *action plan.* This peculiar model leads to the understanding of cultural-emotive dimensions of a student's intense experiences with Gen-AI. It further facilitates an analysis of the students' perspectives and experiences and even categorise such experiences through a demographic structuring. The model lets us understand holistically how culture, value systems, societal framework, and demographic detail affect a subject's interaction with AI-enabled technologies.

Our investigation can claim its relevance and contemporality in view of Gen-AI's pervasive presence within educational domain. A closer analysis of students' voices pertaining to their experiential understanding of opportunities and threats related to AI, is also studied in current research (Chang, Hu, 2023). Such experiential understanding of students' interaction with AI becomes imperative within an UAE context not only given its recent advancement of incorporating AI in education sector, but also, in its chequered multi-cultural demography (Tjondronegoro et al., 2022).

Our study poses three fundamental research questions:

i. Do we perceive an association between students' demographic characteristics and their Gen AI usage frequency?
ii. How do educational background factors relate to students' Gen AI usage patterns?
iii. With regards to Emirati university, how do students experience and perceive Gen AI within the classroom with reference to accessibility, usability, and applicability?

The importance of this mixed-methods research lies in its ability to inform educational policy and practice through demographic analysis and incorporation of reliable student voices. Researchers, (Jewels, Albon, 2011) argue that intrinsic understanding of technological advancement viz. a viz. cultural literacy in UAE higher education demands discernign attention to students' perspective and lived experiences. This perspective becomes imperative in a scenario when aAI technologies are integrated within educational institutes to include voices of the students and specially respond to their cultural backgrounds and educational necessities (Raven, O'Donnell, 2010).

The multicultural demography that impacts the adoption and perception of Gen-AI in UAE classrooms demands examination and critical scrutiny. The element of trust

and reliance on AI by the Emirati University students differs with respect to their socio-cultural backgrounds and subjective understandings and experiences. Further, demographic characteristics such as age, gender, socio-economic factors go to influence a student's access and familiarity of AI within a classroom and broader educational scenario.

Yet, it is important to take into cognizance the broader implication of AI integration. Gen-AI adoption is not only dependant on the technologies own capacity but also the adopters social and institutional standings. For example, resource availability (internet speed, computer infrastructure, and digital platform accessibilities) remains a preconditioned which is fundamental in nature and cannot be dismissed nonchalantly. Resource allocation disparities that constitute digital divide and limits Gen-AI accessibility among students maybe considered a negative force that exacerbates inequalities. Factors beyond resource availabilities that impact Gen-AI applicability for Emirati university students may also be the technologies access usability and applicability elements. A thorough understanding of these factors within a Global South scenarios such as that of UAE brings in the discourses of cultural sensitivity within the educational system designed for Emirati students.

Literature Review

Theoretical Framework (Reflection Model)

The theory behind this study relies on Gibbs' (1988) Reflective Cycle. It looks at how sociocultural and demographic factors impact how simple it is to get, use, and apply AI in the UAE, especially in higher education (Aldosari, 2020). Gibbs' Reflective Cycle is a structured model widely employed in educational settings to facilitate reflective learning. Comprising six stages: *Description, Feelings, Evaluation, Analysis, Conclusion, and Action Plan,* this model guides individuals through a reflective process that aids in understanding their experiences and informing future actions (Campbell, Rogers, 2022). In this paper, the model is applied to examine the interactions of Emirati university students with Gen-AI to provide a structured lens with respect to understanding how peculiar sociocultural and demographic contexts structuralize their experiences. Moreover, the model assisted in guiding the design of reflective questions extended to students and informed the qualitative data analysis.

Using Gibbs' Reflective Cycle (Mollick, Mollick, 2023) as a framework for this study, we can observe how sociocultural and demographic factors influence students' experience with Gen-AI in an organized manner. By navigating students through a reflective process, the model facilitates significant insights into UAE's challenges and opportunities for AI adoption. This method not only assists in comprehending AI's role in several educational settings, but also helps in fostering strategies that align with UAE's vision for incorporating culturally sensitive technological innovation in the field of education.

Sl. No	Name of the Stage	Discussion
1	Description	The first stage of Gibbs' Reflective Cycle-is documenting instances of particular experiences. The students provide their interactive experiences with Gen-AI tools in the classroom setup such as ChatGPT. They outline how AI interacts with others including problematic instances during implementation (Abbas et al., 2023). Such documentation assists in contextualizing their experience within UAE's multicultural setting. The multicultural setup of UAE generates a diverse range of experiences, thereby influencing the nature of accessibility and usability of AI (Najdawi, 2020)
2	Feelings	Provides the students to discuss in detail about their experiences of interacting with Gen-AI tools. Students are free to express their comfort, fears, and excitement about using ChatGPT by expressing his emotions (Alharbi, 2023). This is quite significant within the context of UAE as the multicultural setup may have certain bearing on the emotional responses towards new technologies. For example, students from certain cultural backgrounds may express anxiety while others may reveal a more enthusiastic emotion about integrating AI, which is vital for understanding the subjective nature of their experiences (Kaya et al., 2022)
3	Evaluation	This phase gives learners the opportunity to comment on their experiences with Gen-AI tools such as ChatGPT. Evaluation encourages students to consider positive aspects such as enhanced effectiveness, and assist them in overcoming the challenges of accessing unfamiliar interfaces (Wood, Moss, 2024). Students are encouraged to note how AI applications are used and accessed in the real sense, thereby enabling the parameter to judge where AI has excelled and where potential adjustments could be made to cater to the diverse sociocultural cohorts.
4	Analysis	In this phase, students critically reflect on why certain aspects of their interactions with ChatGPT occurred as they did. This stage is crucial for unmasking how factors such as gender, language proficiency, prior technological experiences, and societal and cultural norms shape their curiosity and engagement with AI (Viberg et al., 2021). The study exposes the complexity of linking specific user experiences to broader societal, cultural, and demographic factors, underscoring the intricate challenges of employing Gen-AI within a multicultural framework (Norouzi et al., 2020).

5	Conclusion	In this stage, students synthesize insights from their earlier stages to identify the lessons learned. The study prompts a critical examination of students' understanding of AI. This helps in uncovering the extent to which Gen-AI tools such as ChatGPT align with or disrupt their cultural expectations. Moreover, this phase encourages students to investigate alternative approaches to engaging with Gen-AI tools, offering valuable insights into the challenges and possibilities of adapting these technologies to accommodate diverse socio-cultural contexts (Zheng et al., 2024)
6	Action Plan	This phase highlights more pragmatic pathways such for adapting ChatGPT skills in being more culturally sensitive and inclusive (Wood, Moss, 2024). As Mollick and Mollick (2023) observed, students suggest ways towards the better use of ChatGPT in a futuristic educational setting and overhauling that schools might take into consideration in making Gen-AI tools relatively more accessible and usable. The ideas that this phase generates leads to futuristic grand goals of transmogrifying AI to socio-culturally sensitive AI, especially within an educational paradigm (Mello et al., 2023)

Accessibility of AI (General in the World and the Gulf Countries, UAE)

Globally, higher education institutions are witnessing rapid integration of Gen AI to support diverse learning environments. Gen AI tools have improved students' learning experiences by offering personalized learning, streamlining research methods, and raising digital literacy (Meakin, 2024). Nevertheless, this advancement in Gen AI in higher education is not without challenges. For example, Walczak and Cellary (2023) warned about the possibility of students' over-reliance on Gen-AI, which could deteriorate their critical and creative thinking capabilities, thereby increasing intellectual laziness. However, as Johnson and Davis (2024) pointed, Gen-AI is influential in bridging gaps for underrepresented students, mainly in the United States of America. Tools such as ChatGPT has allegedly played a vital role in supporting students hailing from disadvantageous backgrounds in negotiating academic hurdles. Nevertheless, paucity of digital infrastructure and insufficient training and skills deter the effective use of Gen AI tools (Johnson, Davis, 2024). In fact, as Walczak and Cellary (2023) points out, developing countries, where technological and infrastructural barriers persist, face these challenges more than the developed nations.

To ensure equitable access to technological advancement across diverse socio-economic groups in the Global South, it is significant to address such AI inequalities. The study reveals significant differences in Gen-AI usage patterns, based on gender, socio-

economic status, and educational background, among Emirati University students. Such disparities reflect a broader systematic issue that hinder the inclusive adoption of AI technologies. By illustrating these disparities, the study underscores the significance of implementing policies and initiatives that aims to bridge the digital segregation. It is essential to foster an inclusive educational environment to ensure that all students, irrespective of their socio-economic background, have access to AI tools and digital literacy skills. Addressing these disparities enhances educational outcomes and promotes social equity and economic development in the Global South.

Despite the constant exponential growth of incorporating Gen-Ai in the Gulf region, the literature revealed that the process of adopting Gen-AI in higher educational institutes still encounters diabolical challenges. This study aims to explore such concerns within the context of UAE. For example, Alotaibi and Alshehri (2023) highlighted some of the issues with Gen-AI accessibility, such as lack of infrastructure, the necessity for teacher training, and the paucity of funds in educational institutes of Saudi Arabia. Similarly, Hail et al. (2024) conducted a study to explore Generation Z's acceptance and adoption of Gen-AI in the higher education institutes of Oman. The study exposed the lack of digital literacy and documented it to be one of the glaring issues that prevents students from benefitting from Gen-AI tools. Such findings reveal the need for a more constructive digital literacy campaigns to facilitate the accessibility to Gen-AI tools.

To counter such challenges, UAE has taken considerable steps to become one of the leading countries in adopting Gen-AI, especially in the education sector. Studies conducted by El Naggar et al. (2024) and Khurma et al. (2023) reveal UAEs efforts to include Gen-AI in regular and inclusive education sector. The transformed education goals of UAE has led to the use of Gen-AI tools to enhance the learning experiences of students with special needs, thereby offering a more customized and scaffolded learning with tailored instructions. Similarly, Bilquise et al. (2024) revealed that ease of use and social influence were significant contributors in motivating students' adoption of Gen-AI tools. Even though the UAE has access to advanced technology, digital literacy gaps and a lack of training still make it hard to use Gen AI technologies effectively in higher education in the UAE (Bilquise et al., 2024).

Moreover, addressing the digital divide issue in the UAE is important (Sallam et al., 2024). Based on a sample of 608 university students in the UAE, 85.4% reported using ChatGPT. However, they were worried about how different groups might be unable to access Gen AI tools because of costs (like needing help to afford them) or infrastructure issues (like needing help to get a reliable internet connection). For example, Sallam et al. (2024) say this can cause differences in Gen AI accessibility.

Usability of AI (Generally in the World and the Gulf Countries, UAE)

Literature on the usability of Gen AI in higher education has pinpointed key factors that determine the effective implementation and usage of Gen AI tools. A study by Sing et al. (2022) investigated secondary school students' usability of AI through the framework of

the Technology Acceptance Model (TAM) and the Theory of Planned Behaviour (TPB). The results revealed that students' perceived ease of use and usefulness were significant predictors of students' intentions to use Gen AI tools. In particular, students tend to adopt Gen AI technologies if they find them user-friendly and helpful in their academic tasks. This result is consistent with the findings from Sanchez-Gonzalez and Terrell (2023), who confirmed that Gen AI-driven classrooms have greatly enhanced students' satisfaction and learning as students appreciated the personalized feedback generated by Gen AI.

Furthermore, empirical studies in the Gulf and Arab regions have shown that students have had positive experiences using Gen AI tools. For instance, Saqr et al. (2024) explored the acceptance and user satisfaction of Gen AI-driven platforms like Blackboard and Moodle among Saudi university students. The results underlined that ease of use and usefulness influence students' satisfaction and attitudes toward Gen AI-driven platforms.

Aburayya et al. (2023) investigated Gen AI usability by exploring the adoption of metaverse systems in UAE universities. The study revealed that perceived ease of use and perceived value are significant predictors of students' intention to adopt these Gen AI tools. Students who saw the metaverse's educational benefits were likelier to use it. This finding is vital in understanding the importance of integrating Gen AI tools that are accessible and intuitive in learning environments to ensure that all students, regardless of their digital skills, can benefit from them.

Applicability of AI (Generally in the World and the Gulf Countries, UAE)

Gen AI in the arena of teaching and learning has significantly shown the potential to improve higher education by incorporating methods that are interactive and delightful therefore making learning more enjoyable for students. For instance, Wood and Moss (2024) studied the impact of Gen AI on students' learning a master's course in National University at Los Angeles. Employing action research in the study validates that combining Gen AI-driven tools with the curriculum acted as a catalyst in students' interest and involvement in studying and has been more aware of the ethical concerns that come up with the inclusion of Gen AI.

In another context, Komba (2024) explored the impact of ChatGPT on students' learning experiences at Mzumbe University. Through qualitative methods, students emphasized that ChatGPT assisted in improving their academic performance, particularly in doing assignments and projects, encouraged self-directed learning, and enhanced their understanding of course materials. However, students insisted that ChatGPT might provide outdated and inaccurate information requiring regular updates.

There are very limited studies done on the applicability of Gen AI in the Gulf and Arab regions, and the existing scholarly works are still in its nascent stages. For instance, Jaboob et al. (2024) examined the adoption of Gen AI tools in Oman, Jordan, and Yemen and the findings underscores the positive impacts of Gen AI tools on the students' cognitive developments and engagements by offering immediate feedback and additionally

providing individual subjective learning instructions. However, the authors have noted that the usage of Gen AI remains limited due to several reasons. The key factors include the lack robust infrastructure in many of the higher educational institutions in the Arab regions, and the resistance caused due to the cultural attitudes toward technology and artificial intelligence, here Gen AI.

In the UAE, there is an increased use of AI into educational purpose and Gen AI is being increasingly integrated into higher educational systems. However, it is still in the early phases of exploration. Through the findings of a study by Mosleh et al. (2024), involving undergraduate students from the Higher College of Technology to understand the relationship between chatbot utilization and emotional intelligence showed that there is significant interconnection between the two. Additionally, results established that students, majorly male and from the STEM disciplines, uses chatbots more than female students. This study enlightens upon the fact that beyond the academic benefits attributed to Gen AI tools, it may play an active role in strengthening students' emotional intelligence, such as self-reflection, emotional regulation, and social awareness.

Literature Gap

While the existing literature highlights Gen AI's accessibility, usability, and applicability in higher educational settings, a significant gap reminds in research based on the demographics of the Gulf countries, particularly the UAE. Although the existing scholarly work explores Gen AI adoption in the Gulf region, they often tend to neglect the socio-cultural factors that is important in shaping its usage patterns. There is lack of comprehensive research on how variables such as gender, high school curriculum, academic performance, and parental education levels influence the accessibility, usability, and accessibility of Gen AI tools within the region's unique socio-cultural and economic contexts that is. This study aims to address this gap by investigating these factors in the UAE's higher education context, providing actionable insights to inform Gen AI policy and practice across the Gulf region.

Based on the literature reviewed, this study hypothesizes the following:

1. *Gender Hypothesis*: Female students exhibit higher frequencies in the weekly usage of Gen AI tools compared to their male counterparts which reflects a gendered preferences in technology adoption and academic engagements.
2. *Age Hypothesis*: Younger students, particularly those in their first and second years of university, are more inclined to use Gen AI tools frequently, reflecting generational familiarity with digital technologies.
3. *School Type Hypothesis*: Students from private or international schools are more likely to use Gen AI tools compared to those from public schools, due to differences in access to technological resources and curricula are focused on digital literacy.
4. *High School Percentage Hypothesis*: Students with higher academic performance (e.g., high school grades above 90%) are more likely to use Gen AI tools everyday

which indicates a correlation between academic achievement and technological adoption.

5. *Program of Study Hypothesis*: Students enrolled in technology centric programs such as Business Transformation or Computational Systems demonstrate higher Gen AI usage compared to those in humanities or arts programs, owning to the nature of their coursework.
6. *Parental Education Hypothesis*: Students whose fathers hold higher educational qualifications (e.g., a bachelor's degree or above) are more likely to exhibit higher daily usage patterns of Gen AI tools, suggesting the influence of familial educational background on technological engagement.

This study probes to validate these hypotheses and provide nuanced insights into the intersection of socio-cultural and demographic factors in relation to Gen AI adoption in the UAE's higher educational sector.

Methodology

Research Design

This study has employed a mixed-method approach, incorporating both the quantitative and the qualitative data collection methods. For the quantitative component, demographic data from students were collected, while in the qualitative component, open-ended reflective responses following Gibbs' reflective cycle were included. This mixed-method approach aids in capturing both the breadth and depth of students' experiences. The approach allows for:

- Statistical analysis of demographic and educational factors influencing AI usage
- Deep exploration of students' personal experiences with AI technologies
- Understanding of cultural and social influences on AI adoption
- Identification of barriers and enablers to AI accessibility and usability
- Recognition of diverse perspectives within the student population

By combining demographic analysis with student voices, this study contributes to the growing body of literature examining the intersection of technology, culture, and education in rapidly developing nations. Furthermore, the mixed-method design of the study allows for a comprehensive exploration of sociocultural and demographic factors that influence student experiences with Gen AI in the classroom at Emirati University. This addresses both the measurable attributes and the subjective perceptions. The insights gained from the study will provide valuable guidance for creating more inclusive and culturally appropriate AI solutions in the higher education sector (Tjondronegoro et al., 2022).

Participants

The sample for the study contained 83 Emirati university students enrolled in three different courses at one university college. The participants, purposefully selected, were

from these three courses, that integrated the ChatGPT into their lesson activities. Participants were selected based on their enrolment in the courses that incorporated Gen AI tools as part of their learning activities. The data was collected during the spring semester of 2024.

Data Collection

The participants completed a two-stage data collection process. First, a demographic survey gathered information about the socio-cultural backgrounds of participants, including age, gender, citizenship status, parental education levels, educational background (high school curriculum), language proficiency, program and year of study, and patterns of Gen AI tool usage including frequency. These variables were selected based on the participant's potential relationship to Gen AI tool adoption in their educational contexts.

In the second stage, based on Gibbs' Reflective Cycle, the participants provided qualitative responses through structured reflections. This framework facilitated a systematic reflection on the experiences of participants with Gen AI tools, encompassing concrete experiences, reflective observations, abstract conceptualization, and plans for future implementation in academic contexts. The structured reflection format enabled consistent data collection across participants while allowing for individual variation in response content and depth.

Data Analysis

A combination of quantitative and qualitative analysis techniques was employed.

Quantitative Analysis

Demographic data were analyzed using descriptive statistics to identify trends in Gen AI accessibility, usability, and applicability among different demographic groups. Comparative analysis was conducted to explore variations based on key demographic variables, such as gender, age, and educational background. To examine the associations and test hypotheses between Gen AI usage frequency and various demographic factors among Emirati university students, chi-square tests of independence were conducted. The analysis included four key variables: father's education level, school type, gender, and high school grade percentage. Chi-square tests were performed to assess the relationships between Gen AI usage frequency (categorized as daily, weekly, monthly, or never) and each of the demographic variables. The strength of associations was evaluated using contingency coefficients. For the analysis involving high school grade percentage, a gamma statistic was also calculated to assess the direction and strength of the ordinal relationship. For all tests, the significance level was set at $p < .05$. Utilizing statistical software Jamovi 28, the analyses were conducted, and the results were deduced based on the chi-square values, degrees of freedom, p-values, and contingency coefficients.

Qualitative Analysis

We implemented a qualitative approach to interpret and analyse the response of Emirati university students to their use and utilization of Gen-AI tools such as ChatGPT within the classroom. The analysis emphasizes on five stages, namely: Feelings, Evaluation, Analysis, Conclusion, and Action Plan of Gibbs' Reflective Cycle (1988). By implementing the reflective model, the study sought to explore and analyse the students' experiences with respect to accessibility, usability, and applicability of ChatGPT in higher education structure.

By employing this reflective model, the study sought to systematically explore and interpret the experiences and perceptions of students regarding accessibility, usability, and applicability of ChatGPT in higher education settings. As a supportive tool during the data analysis, ChatGPT was used to enhance the efficiency and rigor of the process.

Data Collection and Preparation

1. Source of data: 83 Emirati University students who participated in lab sessions integrating ChatGPT into their academic activities were taken into consideration, and their reflections were collected. Open-ended questions aligned with Gibbs' Reflective Cycle were used to gather responses.
2. Data Refining: The responses were critically reviewed to ensure relevance and clarity. Ambiguous or incomplete responses were removed, and duplicate or repetitive entries were consolidated. This was to ensure that only meaningful and distinct reflections were analyzed.
3. Categorization: To facilitate structured analysis, the gathered reflections were organized into the five stages of Gibbs' Reflective Cycle, viz, Feelings, Evaluation, Analysis, Conclusion, and Action Plan.

Data Analysis Process

The analysis was conducted systematically to identify themes, codes, and patterns within the qualitative data. The following steps were undertaken:

1. Familiarization with Data:

- To develop a deep understanding of the content, the responses were read multiple times.
- ChatGPT-aided preliminary summaries of recurring themes were generated, facilitating a dataset overview.

2. Open Coding:

- Each response was then coded by identifying key concepts, ideas, and sentiments expressed by the students. For example, phrases such as "it gave good feedback" and "helped improve my work" were coded as "detailed feedback."

- To expedite the initial coding process, ChatGPT was used in identifying recurring keywords and phrases across the dataset.

3. Categorization into Themes:

- Open codes were grouped into broader, meaningful themes related to the research question: Accessibility, Usability, and Applicability.
- Codes such as "quick responses" and "instant feedback" were categorized under Accessibility, while "detailed feedback" and "clarifying prompts" were placed under Usability.
- ChatGPT was used iteratively to validate and refine these thematic groupings, ensuring alignment with the research objectives.

4. Frequency Analysis:

- The frequency of each code was calculated to identify the most prominent themes and patterns.
- ChatGPT automated the counting process by scanning the dataset for repeated mentions of each code, ensuring accuracy and efficiency.

5. Representative Quote Selection:

- For each code, representative quotes were selected to illustrate the theme and provide qualitative depth to the analysis.
- ChatGPT assisted in extracting concise, relevant quotes by highlighting responses that exemplified specific codes.

6. Integration Across Stages:

- Reflections from the five stages of Gibbs' Reflective Cycle were synthesized to create a cohesive analysis.
- Redundant or overlapping codes were merged, and unique insights from each stage were retained to ensure comprehensive coverage.

7. Interpretation:

- The identified themes were interpreted in the context of the research question. Accessibility focused on the ease of use and immediate feedback provided by ChatGPT, Usability emphasized the interaction between students and the tool, and Applicability examined how students leveraged ChatGPT for academic tasks.
- The final analysis highlighted both the benefits and limitations of using ChatGPT in the classroom.

Role of ChatGPT in Analysis

ChatGPT was employed as a supplementary tool during data analysis. Its contributions included:

1. Identifying patterns and recurring concepts in the dataset.
2. Automating frequency counts for codes, saving time and ensuring accuracy.
3. Assisting in refining thematic groupings and validating the consistency of codes.
4. Extracting representative quotes from large datasets.

5. Providing summaries of preliminary findings, which were reviewed and refined by the researchers.

While ChatGPT streamlined certain aspects of the analysis, all final coding, categorization, and interpretation decisions were made by the researchers to ensure alignment with the study's objectives and the theoretical framework and to assure the accuracy and context-specific interpretation of the data.

Reliability and Validity

To enhance the reliability and validity of the analysis:

1. Triangulation: Reflections were analyzed iteratively by both researchers and ChatGPT to ensure consistent coding and thematic categorization.
2. Peer Review: The thematic structure and interpretations were reviewed by an independent researcher familiar with qualitative methods to validate findings.
3. Participant Authenticity: Representative quotes were directly drawn from the dataset to ensure an authentic representation of students' voices.

Ethical Considerations

Approval of the study was provided by the Social Sciences Research Ethics Committee (ZU24_003_F) of a public university of UAE. To collect the data, we relied on utilizing Google Forms and collated the responses from the students during the Spring and Fall Semesters of 2024. To ensure a multicultural representation, the survey was distributed across multiple sections of general education courses within the university. To capture the comprehensive insights, the survey included both open-ended and close-ended questions. To safeguard the confidentiality, the responses of the students were completely anonymized. All the participants had provided their consent prior to contributing their responses for the survey. The integration of ChatGPT as an analytical tool was disclosed transparently to maintain accountability and trust in the research process. These measures ensured compliance with institutional and academic ethical guidelines throughout the study.

Results

Table 1 provides a detailed demographic profile of the study participants, linking their backgrounds and experiences to their interactions with Gen AI in an academic context.

The findings reveal that most students were Emirati females (68.3%), averaging 19 years old, predominantly enrolled in programs like Business Transformation, Social Innovation, Computational Systems, and Sustainability. Most students had a strong academic background, with 65.1% achieving a high school percentage of 91–100% and 54.4% maintaining a CGPA above 3.0. English accounted for 87.6% of their high school education, indicating a preference for international or private schooling. Nearly all students

(98.4%) reported prior experience with Gen AI, and its frequent usage was evident, with 77% using such tools daily or weekly. The accessibility and perceived value of Gen AI in their academic activities are highlighted by this broad usage. Its application, however, was frequently supplemental, aiding in idea development, content structuring, and language correction. Although they recognised usability issues including the requirement for exact directions and confirming Gen AI-generated comments, they valued its speed and ease. Notwithstanding these drawbacks, Gen AI was well known for its potential to boost academic confidence, encourage creativity, and improve learning results.

Table 1. Demographics of Students Participating in the Study

Characteristic	Category	Count	Percentage
Age	Mean: 19.0 years	-	-
	Range: 18–23 years	-	-
Gender	Female	127	68.3%
	Male	59	31.7%
Citizenship	UAE	174	94.6%
	Other	10	5.4%
Emirate	Abu Dhabi	164	88.2%
	Other Emirates	22	11.8%
High School Type	Private (International)	97	52.2%
	Public	55	29.6%
	Private (UAE Ministry)	28	15.1%
	Other	6	3.1%
Language of Instruction	English	163	87.6%
	Arabic	14	7.5%
	Both/Other	9	4.9%
High School Percentage	91–100%	121	65.1%
	81–90%	56	30.1%
	71–80%	9	4.8%

Program of Study	Business Transformation	58	31.4%
	Social Innovation	50	27.0%
	Computational Systems	41	22.2%
	Sustainability	34	18.4%
	Arts	2	1.1%
Year of Study	Year 1 (2023)	95	51.4%
	Year 2 (2022)	81	43.8%
	Year 3 (2021)	7	3.8%
	Year 4 (2020)	2	1.1%
Current CGPA	3.31–3.70	37	20.3%
	3.01–3.30	32	17.6%
	3.71–4.00	30	16.5%
	2.71–3.00	29	15.9%
	2.31–2.70	22	12.1%
	2.01–2.30	18	9.9%
	Less than 2.00	14	7.7%
Gen AI Usage	Ever used	182	98.4%
	Never used	3	1.6%
Gen AI Frequency	Daily	71	38.2%
	Weekly	72	38.7%
	Monthly	38	20.4%
	Never	5	2.7%

Results of Quantitative Data: Chi-Square Analysis of Associations in Gen AI Usage

To investigate the relationships and test hypotheses between students' demographic traits and their usage habits of Gen AI, we performed a number of chi-square tests of inde-

pendence. Relationships between the frequency of Gen AI usage and four important variables — school type, gender, academic achievement, and father's educational attainment — were examined in the analysis (Table 2).

School Type Hypothesis

It has been observed that pupils attending private international schools are more likely to regularly utilize Gen AI products. The frequency of Gen AI usage and school type were shown to be significantly correlated ($\chi^2(9, N = 186) = 28.9, p < .001$), with a moderate effect size (CC = 0.367). Students from private international schools demonstrated higher daily usage frequencies (O = 40, E = 37.41), while students from public schools reported lower daily usage patterns. This result highlights the impact of school curriculum and technological resources on Gen AI adoption.

Gender Hypothesis

The hypothesis that female students exhibit higher usage of Gen AI tools was confirmed. A statistically significant relationship was observed between gender and Gen AI usage frequency ($\chi^2(3, N = 186) = 12.2, p = .007$), with a smaller effect size (CC = 0.248). Female students showed higher weekly usage frequencies (O = 53, E = 49.16) compared to male students. This finding aligns with gendered differences in technology adoption behavior, particularly in academic settings.

Academic Performance Hypothesis

The hypothesis that students with higher academic performance use Gen AI tools more frequently was partially supported. A significant association was found between high school grade percentage and Gen AI usage frequency ($\chi^2(6, N = 186) = 15.3, p = .018$), with a moderate effect size (CC = 0.276). Students with higher grades (91–100%) exhibited higher daily usage frequencies (O = 50, E = 46.19). However, the gamma statistic ($\gamma = -0.121$, SE = 0.115, 95% CI [−0.346, 0.104]) indicated a weak and non-significant negative trend, suggesting that academic performance may have a minor but inconclusive influence on usage frequency.

Parental Education Hypothesis

The hypothesis that students whose fathers hold higher educational qualifications are more likely to use Gen AI tools frequently was marginally supported. A trend-level association was found between father's education level and Gen AI usage frequency ($\chi^2(15, N = 186) = 24.9, p = .051$), with a moderate effect size (CC = 0.344). Students whose fathers held bachelor's degrees reported higher daily usage patterns (O = 41, E = 32.45), indicating a potential influence of family education background on technology adoption.

Age and Program of Study Hypotheses

Contrary to expectations, no significant associations were found between Gen AI usage frequency and students' age or program of study. These variables, while relevant in prior studies, did not demonstrate a significant impact in the current sample, as summarized in Table 2.

Table 2. Chi-Square Tests of Independence Results for Associations with Gen AI Usage Frequency (N = 186)

Variable	Hypothesis	χ^2	df	p	CC	Usage Pattern Notes
H3:School Type	Students from private international schools use Gen AI tools more frequently.	28.9	9	< .001**	0.367	Higher daily usage in private international schools (O = 40, E = 37.41)
H1:Gender	Female students exhibit higher usage of Gen AI tools compared to male students.	12.2	3	.007**	0.248	Female students showed higher weekly usage (O = 53, E = 49.16)
H3:High School Percentage	Students with higher academic performance (91–100%) use Gen AI tools more frequently.	15.3	6	.018*	0.276	Students with 91-100% grades showed higher daily usage (O = 50, E = 46.19)
H6: Parents' Education	Students with fathers holding higher education degrees are more likely to use Gen AI tools.	24.9	15	.051†	0.344	Bachelor's degree holders showed higher daily usage (O = 41, E = 32.45)
H2:Age	Younger students (first and second year) use Gen AI tools more frequently.	N/A	N/A	N/A	N/A	No significant association observed
H5:Program of Study	Students in technology-centric programs use Gen AI tools more frequently than humanities students.	N/A	N/A	N/A	N/A	No significant association observed

Note: CC = Contingency Coefficient; O = Observed frequency; E = Expected frequency †p < .10, *p < .05, **p < .01

Qualitative Data Results: Thematic Analysis

This study utilized Gibbs' Reflective Cycle (1988) to explore how Emirati university students experience and perceive Gen AI in the classroom. Five stages of the reflective cycle — Feelings, Evaluation, Analysis, Conclusion, and Action Plan — were analyzed qualitatively to answer the research question, "How do Emirati university students experience and perceive Gen AI in the classroom regarding accessibility, usability, and applicability?" We present the data analysis below, synthesizing themes and codes across all stages and their interpretation.

Table 3. Consolidated Analysis of Five Stages of Gibbs' Reflective Cycle (1988).

Theme	Code	Frequency	Representative Quotes	Interpretation
Accessibility	Instant Access	18	«Provides instant feedback on what I'm looking for.» «Quick responses save time when working on assignments.»	ChatGPT's instant responses made it highly accessible, providing students with immediate feedback for academic tasks.
	Time Efficiency	14	«Helped save time by quickly addressing my questions.» «A good substitute when I'm in a hurry.»	Students valued ChatGPT's efficiency, especially in time-constrained situations, enhancing its accessibility for quick support.
	Supplementary Tool	12	«Useful for getting extra feedback alongside the professors.» «Good as an additional perspective to refine my work.»	People viewed ChatGPT as an additional resource that complemented traditional feedback forms rather than completely replacing them.
Usability	Detailed Feedback	24	«Gives specific feedback to improve my work.» «Helped refine parts I hadn't noticed before.»	ChatGPT was praised for providing detailed, actionable feedback that helped students identify areas for improvement.
	Instruction Sensitivity	15	«You need to give clear instructions for accurate feedback.» «I had to refine my prompts multiple times for better results.»	Effective use of ChatGPT required clear and specific prompts, reflecting a usability challenge that students needed to navigate.
	Verification of Responses	13	«I double-check its suggestions with the professor.» «Not all feedback aligns with assignment requirements.»	Students often verified ChatGPT's output, demonstrating moderate trust in its accuracy and alignment with academic expectations.

	Simplified Explanations	10	«Breaks down complicated ideas into simpler terms.» «Helped clarify concepts I didn't understand well.»	ChatGPT was appreciated for its ability to simplify complex topics and support comprehension and usability.
	Feedback Diversity	11	«Combining AI and peer feedback gave me better insights.» «Both professor and ChatGPT feedback were useful for improvements.»	Students valued a combination of ChatGPT and human feedback, using diverse sources to enhance their understanding and confidence.
Applicability	Content and Idea Generation	18	«It gave me ideas I hadn't thought of.» «Helped brainstorm creative approaches to my assignment.»	ChatGPT was widely applied as a brainstorming tool, supporting creativity and helping students generate new ideas for assignments.
	Grammar and Language Support	16	«Helps with grammar and professional language.» «Fixes grammar mistakes and makes writing clearer.»	Students used ChatGPT to refine grammar and improve language clarity, enhancing their overall writing quality.
	Task-Specific Refinements	10	«I could have given ChatGPT the rubric for more targeted feedback.» «Its feedback was too general for some assignments.»	We noted ChatGPT's limited ability to address assignment-specific requirements, suggesting the need for more tailored interactions.
	Structuring and Organization	15	«Helps organize paragraphs logically.» «Made it easier to structure my ideas effectively.»	Students appreciated ChatGPT's ability to enhance the organization and structure of their work, making it applicable to academic tasks.
	Academic Confidence Building	8	«Gives me confidence before submitting my work.» «Helped validate my ideas and writing approach.»	ChatGPT provided reassurance by validating students' work and boosting their confidence in their submissions.

Qualitative analysis of student reflections through Gibbs' Reflective Cycle framework revealed three predominant themes: Accessibility, Usability, and Applicability. These themes emerged as central to understanding Emirati university students' experiences and perceptions of Gen AI integration in educational contexts.

Accessibility

Analysis indicated that participants (N = 83) consistently emphasized ChatGPT's accessibility characteristics. Students reported that the tool's capacity for instantaneous feedback facilitated more efficient academic task management. The data suggested that participants viewed ChatGPT as a complementary resource to traditional feedback mechanisms (e.g., instructor feedback, peer review) rather than a replacement. One participant noted, "It gave us good feedback that will be useful for our future writing," exemplifying the tool's perceived accessibility for immediate academic support.

Usability

The usability theme encompassed both the advantages and limitations of ChatGPT implementation. Participants frequently cited the tool's capability to provide detailed feedback and clarify complex concepts. However, the data revealed that effective use required accurate and prompt engineering, and students needed to check their work against assignment criteria regularly. The study found a pattern of strategic integration in which users combined ChatGPT feedback with more traditional academic support systems to make them easier to use. As one student articulated, "Professor's feedback is better," highlighting the complementary nature of Gen AI-assisted learning.

Applicability

Examination of student reflections revealed diverse applications of ChatGPT across academic tasks. Primary applications included ideation processes, grammatical refinement, structural improvements, and organizational enhancement of assignments. However, participants consistently reported limitations regarding assignment-specific criteria alignment. The data suggested a correlation between ChatGPT usage and increased academic self-efficacy, with students reporting enhanced confidence in their work following Gen AI-assisted review. One participant observed that it "helps me understand many things," indicating the tool's broad applicability while acknowledging its constraints.

Synthesis of Findings

The study found that Emirati college students see ChatGPT as an easy-to-use academic tool that improves learning and productivity. However, its usefulness depends on how well users know how to use prompts and check their work. Participants liked the features that gave them immediate feedback and helped them organize. However, the data showed two main problems: (a) they were not perfectly aligned with the assignment requirements, and (b) the best way to interact with them depended on the user's knowledge. These findings suggest that Gen AI functions optimally as a supplementary academic tool rather than a standalone resource. The results indicate that systematic integration of

prompt engineering and critical evaluation skills into academic curricula may be necessary to optimize educational outcomes. This could make it easier for students to use Gen AI tools strategically in their learning.

Discussion

Demographic and Educational Factors Shaping Gen AI Usage

This study confirmed the hypothesis that socio-cultural factors, such as school type and gender, significantly influence Gen AI adoption. The results demonstrate that demographic and educational factors, including the type of high school attended, gender, and high school grade percentage, significantly influence Gen AI use patterns. A slightly significant trend was also observed for the father's education level, as shown in Table 2. The findings draw attention to the influence of students' educational backgrounds and academic performance on their engagement with Gen AI technologies.

Research shows that Gen AI adoption is higher in schools with access to advanced technology and digital literacy education (Alotaibi & Alshehri, 2023; Sallam et al., 2024). The strong ties to private and international high schools align with this. For example, Table 2 shows that students from private international schools use Gen AI more daily. This shows that educational resources do affect adoption patterns. Gender differences in Gen AI use, with women using it more often each week (Table 2), align with larger trends in how people adopt new technologies. Differences in academic interests and familiarity with digital tools could explain this pattern (Sing et al., 2022). However, the result of gender differences in this study differs from previous studies, which reported that male students used chatbots more than females (Mosleh et al., 2024).

These results enhance previous research by providing a clearer understanding of the impact of sociodemographic and educational factors on millennials' use of AI, particularly in the UAE, where traditional customs and rapid technological advancements coexist. By looking into these factors, the study shows how demographics (like gender or level of education) affect students' engagement with Gen AI. The study also provides educators and policymakers with valuable recommendations to guarantee equitable access to Gen AI tools for all students and their effective integration into educational practices.

Accessibility of Gen AI

The findings of this study highlighted accessibility to Gen AI tools as a central theme (Table 3). Students repeatedly highlighted that ChatGPT provides immediate feedback and saves their time effectively, which reflects its ease of access and utility in academic contexts. These findings support previous research suggesting that instant responses and time-saving features of Gen AI improved students' academic experiences (Meakin, 2024). In particular, students' widespread usage of Gen AI (nearly 98%) (Table 2) emphasizes the accessibility of Gen AI tools.

Gibbs' Reflective Cycle, especially the Description and Evaluation stages, shed light on how students felt about Gen AI's potential to help them with schoolwork in addition to traditional methods (Table 3). Khurma et al. (2023) have documented similar conclusions, acknowledging that Gen AI can provide personalized learning for students and improve their access to educational resources, especially in culturally diverse settings like the UAE. However, this study adds to the existing literature by highlighting the persistent disparities in digital literacy and infrastructural challenges, even in technologically advanced settings like the UAE, despite the general perception of Gen AI as accessible (Sallam et al., 2024).

By looking at more than just the availability of technical tools, these results fill a gap in the research on Gen AI accessibility. Instead, they explore accessibility by students' perceived ease of use in their daily academic tasks (Table 3). They underscore the necessity for focused training programs to promote digital literacy and ensure the effective leverage of Gen AI tools across diverse student populations.

Usability of Gen AI

The study results show that one of the main themes is how useful Gen AI tools are. Students liked how tools like ChatGPT gave them detailed feedback, made complicated ideas easier to understand, and gave them feedback from various sources (Table 3). On the other hand, students reported challenges like the need for instruction sensitivity and verification of responses, which underscore the importance of prompt engineering and critical evaluation. The Technology Acceptance Model (Sing et al., 2022) emphasizes that perceived usefulness and ease of use are key factors in technology adoption (Aburayya et al., 2023; Saqr et al., 2024), and these results corroborate these ideas. Through the lens of Gibbs' Reflective Cycle, the Analysis and Action Plan stages demonstrated how students navigated the Gen AI tool's limitations by complementing Gen AI responses with instructors' feedback to enhance their academic performance (Table 3). This strategic combination corresponds to previous studies highlighting the importance of integrating AI-generated feedback with human input to boost usability in educational contexts (Escalante et al., 2023; Sanchez-Gonzalez, Terrell., 2023). By examining Gen AI usability issues specific to the UAE, such as the need for clear prompts and the level of trust in AI-generated results, these results contribute to the existing literature. It stresses the importance of integrating Gen AI literacy and critical evaluation skills into educational programs to empower students to navigate Gen AI tools confidently. This result fills in a significant gap in the research by looking into how people and AI interact in educational settings with many different cultures.

Applicability of Gen AI

Students agreed that Gen AI tools could help create content, improve language, and organize schoolwork. This showed how versatile tools like ChatGPT support different learn-

ing areas (Table 3). Previous research has shown that including Gen AI in the curriculum improves students' creative and organized ability and interest in learning (Jaboob et al., 2024; Wood, Moss, 2024). These results back this up.

The Conclusion and Action Plan stages of Gibbs' Reflective Cycle showed that students saw Gen AI like ChatGPT as a helpful tool that helped them feel more confident in school and encouraged self-directed learning, especially when it came to coming up with new ideas and improving their grammar (Table 3). Still, the results showed that Gen AI has some problems that make it less valuable, like the inability to meet assignment needs and specific details without full user customization. This is consistent with study by Mosleh et al. (2024), which highlights the need to integrate user training and specialized rubrics.

Conclusion

The study have wider implications for the Global South, particularly in comprehending the social and demographic aspects that impacts the application of Gen AI in higher education. There is considerable correlation between the use of Gen AI and variables including school type, academic success, and gender. These insights can create policies and programs that address unique difficulties faced by students in the Global South, such as infrastructural limitations and gaps in digital literacy. The study's emphasis on the cultural dimensions of AI adoption provides a framework for examining how social norms and cultural values impact technology use. This understanding creates an essentially cultural relevant AI solutions that satisfy the needs of diverse student populations in the Global South.

In Global south Policymakers play a crucial role in shaping and redesigning of AI adaptation. The study provide valuable insights for promoting the equitable and practical application of Gen AI methods in higher education. Investing in digital infrastructure that ensures easy internet access is the first step in building a digital ecosystem. Second, comprehensive digital literacy initiatives to teach children how to use AI technology must be implemented. Third, promoting cooperation between academic institutions, IT firms, and lawmakers can develop relevant AI solutions. Lastly, to consider the needs and values of diverse student groups in the Global South while incorporating AI. By incorporating moral guidelines and legal frameworks we can solve bias, accountability, and transparency issues in AI systems.

Moreover, adopting relevant AI strategies are crucial for effectively integrating AI technology into Global South educational context. This study highlights the significance of implementing AI solutions to students' distinct social and demographic traits in the UAE. Considering social norms and cultural values politicians and educators may increase the efficacy and acceptability of AI technology. For example, improve accessibility and usability by developing multilingual and culturally sensitive AI systems. Including local communities in the development and evaluation of AI technologies helps to ensure the requirements and expectations of users. Our research supports an inclusive approach

to AI adoption that celebrates and honors the diversity of the Global South by prioritizing cultural sensitivity.

Limitations and Future Directions

Several methodological and contextual limitations warrant consideration when interpreting the findings of this study. First, the data collection process presents notable constraints. We obtained the reflections within a structured academic environment, which could introduce social desirability bias in student responses. Also, the short, one-point-in-time nature of the responses may only show a small part of how complicated and changing students' experiences with ChatGPT in school are. The chi-square tests employed in this study only demonstrate associations between variables and cannot establish causal relationships. Sample characteristics impose additional constraints on the generalizability of findings.

The study's focus on Emirati university students in a specific institutional context limits the transferability of results to other cultural and educational settings. Variation in participants' English language proficiency may have influenced their ability to fully articulate their experiences, potentially affecting the richness of the qualitative data. The analysis represents a snapshot during the early adoption phase of ChatGPT in educational settings. With Gen AI technology and ways of using it in the classroom changing so quickly, these results might only partially show how students use Gen AI tools over the long term. The ChatGPT rules at that school may have affected how students thought about and used it differently than in other schools. Despite the drawbacks, this study offers a fundamental insight into how Emirati university students perceive and use Gen AI tools.

Integrating AI system in the Global South raises significant concerns in ethical dilemmas and cultural prejudices. This study demonstrates how sociocultural factors influence students' use of Gen AI technologies. It also illustrates how prejudice AI systems in perpetuate inequity. For example, fundamental trust and dependability difficulties are reflected in the requirement for accurate prompts and the verification of AI-generated outputs. These difficulties highlight how culturally aware AI systems is, and crucial creating transparent, responsible. A cooperative method should address these ethical issues, including educators, legislators, and computer developers. This study adds to the larger discourse on developing AI technologies that are equitable, inclusive, and considerate of cultural variety by promoting discussion on the moral practices of AI adoption.

The results investigate and points several directions for further research. Employing mixed-methods longitudinal studies can offer more insights into students' changing attitudes and usage habits. More sophisticated statistical techniques should be used to eliminate potentially misleading variables and investigate the interactions between academic success, the usage of Gen AI technologies, and other significant aspects. Researches that compare cultures might shed light on how various educational settings affect the adoption and use of Gen AI tools. Finally, we can identify best practices for using ChatGPT in

schools if we closely examine the relationship between the sorts of students, their prior experience with Gen AI, and the tool's functionality. Understanding how Gen AI technologies integrate into various academic and cultural contexts will be essential to fostering equitable and prosperous adoption as these technologies continue revolutionizing education.

This study probes to validate these hypotheses and provide nuanced insights into the intersection of socio-cultural and demographic factors in relation to Gen AI adoption in the UAE's higher educational sector.

References

Abbas N., Ali I., Manzoor R., Hussain T., AL Hussaini M. H. (2023) Role of artificial intelligence tools in enhancing students' educational performance at higher levels. *Journal of Artificial Intelligence, Machine Learning, and Neural Network,* vol. 3, no 5, pp. 36–49.

Aburayya A., Salloum S. A., Alderbashi K. Y., Shwedeh F., Shaalan Y., Alfaisal R., Malaka S. J. M., Shaalan K. (2023) SEM-machine learning-based model for perusing the adoption of metaverse in higher education in UAE. *International Journal of Data and Network Science,* vol. 7, no 2, pp. 667–676.

Aldosari S. A. M. (2020) The future of higher education in the light of artificial intelligence transformations. *International Journal of Higher Education,* vol. 9, no 3, pp. 145–150.

Alharbi L. A. (2023) A systematic literature review on AI algorithms and techniques adopted by e-learning platforms for psychological and emotional states. *International Journal of Advanced Computer Science and Applications,* vol. 14, no 2.

Alotaibi N. S., Alshehri A. H. (2023) Prospers and obstacles in using artificial intelligence in Saudi Arabia higher education institutions — The potential of AI-based learning outcomes. *Sustainability,* vol. 15, no 13, p. 10723.

Bamatraf S., Amouri L., El-Haggar N., Moneer A. (2021) Exploring the socio-economic implications of artificial intelligence from higher education students' perspective. *International Journal of Advanced Computer Science and Applications,* vol. 12, no 6, pp. 369–376.

Ben Romdhane S., Lee S., Al-Shaebi S. (2023) Enhancing Sustainability Communication among UAE's Higher Education Students: The Relationship between Sustainable Living Knowledge and Intention to Live Sustainably. *Sustainability,* vol. 15, no 15, p. 11892.

Bilquise G., Ibrahim S., Salhieh S. M. (2024) Investigating student acceptance of an academic advising chatbot in higher education institutions. *Education and Information Technologies,* vol. 29, no 3, pp. 6357–6382.

Campbell F., Rogers H. (2022) Through the looking glass: A review of the literature surrounding reflective practice in dentistry. *British Dental Journal,* vol. 232, no 10, pp. 729–734.

Chan C. K. Y., Hu W. (2023) Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education,* vol. 20, no 43.

El Naggar A., Gaad E., Inocencio S. A. M. (2024) Enhancing inclusive education in the UAE: Integrating AI for diverse learning needs. *Research in Developmental Disabilities,* no 147, p. 104685.

Escalante J., Pack A., Barrett A. (2023) AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, vol. *20,* no 1, pp. 57–20.

Gibbs G. (1988) *Learning by doing: A guide to teaching and learning methods,* Oxford Polytechnic.

Hail G. A. M. T., Yusof S. A. M., Rashid A., El-Shekeil I., Lutfi A. (2024) Exploring factors influencing Gen Z's acceptance and adoption of AI and cloud-based applications and tools in academic attainment. *Emerging Science Journal,* vol. 8, no 3, pp. 815–836.

Jaboob M., Hazaimeh M., Al-Ansi A. M. (2024) Integration of Gen AI techniques and applications in student behavior and cognitive achievement in Arab higher education. *International Journal of Human–Computer Interaction,* pp.1–14.

Jewels T., Albon R. (2011) Reconciling culture and digital literacy in the United Arab Emirates. *International Journal of Digital Literacy and Digital Competence,* vol. 2, no 2, pp. 27-39.

Johnson D., Alsharid M., El-Bouri R., Mehdi N., Shamout F., Szenicer A., Toman D., Binghalib S. (2022) An experience report of executive-level artificial intelligence education in the United Arab Emirates. *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 36, no 11, pp. 12766–12773.

Johnson D. L., Davis C. G. (2024) Bridging the gap for underserved populations: Personalized AI solutions for college access and learning support. *New Directions for Higher Education,* no 207, pp. 47–62.

Kaya F., Aydin F., Schepman A., Rodway P., Yetişensoy O., Demir Kaya M. (2022) The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence. *International Journal of Human–Computer Interaction,* vol. 40, no 2, pp. 497–514.

Khurma O. A., Ali N., Hashem R. (2023) Critical reflections on ChatGPT in UAE education: Navigating equity and governance for safe and effective use. *International Journal of Emerging Technologies in Learning (iJET),* vol. 18, no 14, pp. 188–199.

Komba M. M. (2024) The influence of ChatGPT on digital learning: Experience among university students. *Global Knowledge, Memory and Communication*, vol. 17, no 2, pp. 152-167.

Lee S., Al-Shaebi S. (2023) Enhancing Sustainability Communication among UAE's Higher Education Students: The Relationship between Sustainable Living Knowledge and Intention to Live Sustainably. *Sustainability,* vol. 15, no 15, p. 11892.

Meakin L. (2024) Exploring the Impact of Generative Artificial Intelligence on Higher Education Students' Utilization of Library Resources. *Information Technology and Libraries,* vol. 43, no 3, pp. 1–13.

Mello R. F., Freitas E., Pereira F. D., Cabral L., Tedesco P., Ramalho G. (2023) Education in the age of Gen AI: Context and Recent Developments. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2309.12332

Mollick E. R., Mollick L. (2023) Assigning AI: Seven approaches for students, with prompts. *SSRN*. https://doi.org/10.2139/ssrn.4475995

Mosleh S. M., Alsaadi F. A., Alnaqbi F. K., Alkhzaimi M. A., Alnaqbi S. W., Alsereidi W. M. (2024) Examining the association between emotional intelligence and chatbot utilization in education: A cross-sectional examination of undergraduate students in the UAE. *Heliyon*, vol. 10, no 11, p. e31952.

Najdawi A. (2020) Assessing AI readiness across organizations: The case of UAE. *Proceedings of the 11th ICCCNT,* pp. 1–7.

Norouzi N., Chaturvedi S., Rutledge M. (2020) Lessons learned from teaching machine learning and natural language processing to high school students. *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 34, no 09, pp. 13397–13403.

Raven J., O'Donnell K. (2010) Using digital storytelling to build a sense of national identity amongst Emirati students. *Education, Business and Society: Contemporary Middle Eastern Issues,* vol. 3, no 3, pp. 201–217.

Sallam M., Elsayed W., Al-Shorbagy M., Barakat M., El Khatib S., Ghach W., Alwan N., Hallit S., Malaeb D. (2024) ChatGPT usage and attitudes are driven by perceptions of usefulness, ease of use, risks, and psycho-social impact: A study among university students in the UAE. *Frontiers in Education,* vol. 9, Article 1414758.

Sanchez-Gonzalez M., Terrell M. (2023) Flipped classroom with artificial intelligence: Educational effectiveness of combining voice-over presentations and AI. *Cureus,* vol. 15, no 11, p. e48354.

Saqr R. R., Al-Somali S. A., Sarhan M. Y. (2024) Exploring the acceptance and user satisfaction of AI-driven e-learning platforms (Blackboard, Moodle, Edmodo, Coursera, and edX): An integrated technology model. *Sustainability,* vol. 16, no 1, p. 204.

Shamout F. E., Ali D. A. (2021) The strategic pursuit of artificial intelligence in the United Arab Emirates. *Communications of the ACM*, vol. 64, no 4, pp. 57-59.

Sing C. C., Teo T., Huang F., Chiu T. K. F., Xing wei Y. (2022) Secondary school students' intentions to learn AI: Testing moderation effects of readiness, social good and optimism. *Educational Technology Research and Development,* vol. 70, no 3, pp. 765–782.

Tjondronegoro D., Yuwono E., Richards B., Green D., Hatakka S. (2022) Responsible AI implementation: A human-centered framework for accelerating the innovation process. *arXiv.* https://doi.org/10.48550/arXiv.2209.07076

Viberg O., Cukurova M., Feldman-Maggor Y., Alexandron G., Shirai S., Kanemune S., Wasson B., Tømte C., Spikol D., Milrad M., Coelho R., Kizilcec R. F. (2023) Teachers' trust and perceptions of AI in education: The role of culture and AI self-efficacy in six countries. *arXiv*. https://discovery.ucl.ac.uk/id/eprint/10184606/

Walczak K., Cellary W. (2023) Challenges for higher education in the era of widespread access to generative AI. *Economics and Business Review,* vol. 9, no 2, pp. 71–100.

Wood D., Moss S. H. (2024) Evaluating the impact of students' Gen AI use in educational contexts. *Journal of Research in Innovative Teaching & Learning*. https://doi.org/10.1108/jrit-06-2024-0151

Yusuf A., Pervin N., Román-González M. (2024) Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education,* vol. 21, no 21.

Zheng C., Yuan K., Guo B., Mogavi R. H., Peng Z., Ma S., Ma X. (2024) Charting the future of AI in project-based learning: A co-design exploration with students. *Proceedings of the CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613904.3642807

Влияние социокультурных и демографических факторов на доступность, удобство использования и применимость генеративного искусственного интеллекта в ОАЭ

Айжан Шомотова
Колледж междисциплинарных исследований
Университет Зайда
Адрес: почтовый ящик 144534, Абу-Даби, ОАЭ
E-mail: aizhan.shomotova@zu.ac.ae

Салва Хуссейн
Колледж междисциплинарных исследований
Университет Зайда
Адрес: почтовый ящик 144534, Абу-Даби, ОАЭ
E-mail: salwa.husain@zu.ac.ae

Аридж Эльсаяри
Колледж междисциплинарных исследований
Университет Зайда
Адрес: почтовый ящик 144534 Абу-Даби, ОАЭ
E-mail: areej.elsayary@zu.ac.ae

Ариндам Дас
Профессор Школы свободных искусств Университета Аллайанс
Адрес: Чиккахагаде-Кросс, главная дорога Чандапура-Анекал
Анекал, Бангалор — 562106, Карнатака, Индия
E-mail: arindam.das@alliance.edu.in

Гленн Масхерт
Профессор Факультета общественного здравоохранения и эпидемиологии в Колледже лечебного дела и медицинских наук Университета науки и технологий в Халифе
Адрес: Почтовый ящик 127788, Абу-Даби, ОАЭ
E-mail: glenn.muschert@ku.ac.ae

Несмотря на стремительную интеграцию искусственного интеллекта во всемирное образование, роль социокультурных и демографических факторов в этом процессе в таких культурно разнообразных странах как, например, ОАЭ остается недостаточно изученной. В данном исследовании изучаются доступность, удобство использования и применимость инструментов генеративного ИИ (Gen AI), таких как ChatGPT, в высшем образовании на примере студентов

из университетов в Эмиратах. Используя смешанную исследовательскую методологию, с помощью демографических исследований и структурированных размышлений, базирующихся на рефлексивном цикле Гиббса, были получены данные 83 студентов. Количественный анализ показал тесную связь между использованием генеративного искусственного интеллекта и такими демографическими и образовательными факторами, как тип старшей школы, гендер и академическая успеваемость, сочетающимися с некоторой зависимостью от отцовского образовательного уровня. Учащиеся международных частных школ сообщили о более частом повседневном использовании ИИ, в то время как женщины делали это более регулярно, что подчеркивает влияние образовательного уровня и гендерных различий на внедрение новых технологий. Тематический анализ выявил в качестве ключевой темы проблему общей доступности. При этом учащиеся подчеркивали преимущества мгновенной обратной связи и экономии времени при использовании инструментов генеративного ИИ. Исследование простоты и удобства использования генеративного искусственного интеллекта показало, что студентам нравится насыщенная обратная связь и то, как ИИ может просто излагать сложные идеи. Однако они столкнулись с такими проблемами как чувствительность запроса и необходимость проверки конечного результата. Результаты проверки общей применимости генеративного ИИ продемонстрировали помощь ChatGPT в создании контента, улучшении языка и организации учебного материала, но также вывили сложность следования отдельным правилам конкретных заданий. Данное исследование вносит вклад в литературу по теме, рассматривая социально-демографические факторы, влияющие на внедрение генеративного ИИ в такой культурно разнообразной среде, как ОАЭ. Результаты показывают, что навыки ИИ -грамотности и критической оценки искусственного интеллекта должны преподаваться в школах, а также необходима стратегия обеспечения всеобщего и равного доступа к ИИ-инструментам. Рекомендации направлены на помощь преподавателям в использовании потенциала генеративного ИИ, смягчая при этом его вызовы для высшего образования.

Ключевые слова: генеративный ИИ, социально-демографические показатели, образование, ОАЭ, использование, доступность

Appendix: Detailed Qualitative Analysis for All Stages of Gibbs' Reflective Cycle

This appendix presents a detailed analysis of each stage of Gibbs' Reflective Cycle — Feelings, Evaluation, Analysis, Conclusion, and Action Plan — based on student reflections. Themes, codes, frequencies, representative quotes, and interpretations are outlined for each stage.

1. Feelings

Theme	Code	Frequency	Representative Quotes	Interpretation
Emotional Response	Positive Experience	18	«It was fun.» «I feel good; it helped develop our work.»	Students expressed positive emotions, associating ChatGPT with enjoyment and learning enhancement.
	Discomfort or Uncertainty	14	«It feels weird and odd.» «I found it unnatural using ChatGPT during the session.»	Some students felt discomfort or skepticism about integrating AI into academic tasks, reflecting unfamiliarity with the tool.

	Ambivalence	16	«Weird feelings, kind of ironic but still fun.»	Students exhibited mixed feelings, recognizing both the novelty and utility of ChatGPT.
Trust and Reliability	Skepticism	12	«It helps but I don't trust it.» «I feel less confident about the answers retrieved from ChatGPT.»	Skepticism arose from concerns about reliability and trustworthiness, limiting students' confidence in fully relying on ChatGPT.
	Human Preference	10	«Professor feedback is better.» «Getting peer-reviewed was more effective.»	Students valued human feedback over ChatGPT, preferring the guidance of professors and peers for critical academic evaluations.

2. Evaluation

Theme	Code	Frequency	Representative Quotes	Interpretation
Feedback Quality	Detailed Feedback	23	«It gave good and specific feedback.» «It highlighted what I was missing in my introduction.»	Students appreciated ChatGPT's detailed feedback, which was seen as helpful for identifying areas for improvement.
	General Feedback	15	«Provides a general idea about what my prompt is.» «Gives good general feedback for improving work.»	General feedback was valued for providing an overview of students' work, even if it lacked specific alignment with assignment needs.
Accuracy Concerns	Reliability Issues	18	«Helpful for feedback but often inaccurate.» «Feedback can be wrong or misleading.»	Students noted accuracy issues, highlighting the need to verify ChatGPT's suggestions.
Ease of Use	Instant Feedback	12	«Gives you instant feedback on what you're looking for.» «Fast answers made understanding easier.»	Instant responses were a major benefit, enhancing ChatGPT's usability for time-sensitive academic tasks.
	Complexity in Output	11	«Sometimes provides complex suggestions that don't help.» «Overly complicated feedback made it hard to follow.»	Overly complex or advanced responses were sometimes unhelpful, reducing the usability of ChatGPT for specific academic contexts.

3. Analysis

Theme	Code	Frequency	Representative Quotes	Interpretation
Learning Impact	Content Improvement	18	«Helped me see where to improve.» «Gave me ideas for improvement.»	Students recognized ChatGPT's role in identifying weaknesses and suggesting ways to enhance their assignments.
	Idea Generation	14	«Gave me ideas I hadn't thought of.» «Helped brainstorm new approaches to my work.»	Idea generation was a significant benefit, supporting creativity and exploration in academic tasks.
	Skill Development	17	«Helped improve grammar and refine my writing.» «Makes my work look more professional.»	ChatGPT contributed to skill development, particularly in writing, grammar, and organization.
Task-Specific Challenges	Generalized Responses	12	«Its feedback doesn't always match assignment requirements.» «I could have given it the rubric for better feedback.»	Generalized feedback sometimes limited ChatGPT's utility for task-specific academic needs.

4. Conclusion

Theme	Code	Frequency	Representative Quotes	Interpretation
Feedback Strategies	Seeking Multiple Perspectives	14	"I could have sought feedback from ChatGPT, peers, and my professor.»	Students recognized the value of combining AI-generated and human feedback for a well-rounded evaluation.
	Refining Prompts	15	«I could have been more detailed with my questions.» «Providing efficient prompts would give better results.»	Prompt refinement was seen as an area for improvement, highlighting the need for effective communication with ChatGPT.
Time Management	Iterative Feedback Use	12	«I would give ChatGPT more time to check multiple versions of my work.»	Students suggested using ChatGPT iteratively for progressive improvements, reflecting a strategic approach to feedback application.

5. Action Plan

Theme	Code	Frequency	Representative Quotes	Interpretation
Future Applications	Feedback Utilization	22	«I will use ChatGPT to get feedback and refine my work.» «I'll use it for structuring paragraphs and improving ideas.»	Students planned to apply ChatGPT's feedback to enhance writing quality and organization in future assignments.
	Prompt Optimization	15	«I'll write well-structured prompts for better responses.» «Providing detailed questions will help improve its accuracy.»	Students aimed to refine their prompts to optimize ChatGPT's output and align it with their academic needs.
	Independent Learning Integration	12	«I will use it alongside professor feedback.» «I'll apply its feedback but verify with other sources.»	Students emphasized using ChatGPT as a supplementary tool, integrating its outputs with independent learning and instructor input.

doi: 10.17323/1728-192x-2024-4-173-195

AI Impacts, Concerns, and Perspectives in the Global South A Thought Leadership Round Table

Arindam Das
Professor, School of Liberal Arts
Alliance University
Address: Chikkahagade Cross, Chandapura-Anekal Main Road
Anekal, Bangalore — 562106, Karnataka, India
E-mail: arindam.das@alliance.edu.in

Glenn Muschert
Professor, Department of Public Health and Epidemiology
College of Medicine and Health Sciences
Khalifa University or Science and Technology
Address: PO Box 127788 Abu Dhabi, UAE
E-mail: glenn.muschert@ku.ac.ae

Mohan J. Dutta
Dean's Chair Professor of Communication
Director, CARE, School of Communication, Journalism, and Marketing
Massey University
Address: Private Bag 11 222 Palmerston North 4442 New Zealand
E-mail: mohanjdutt@gmail.com

Muhammed Bilgehan Aytaç
Assistant Professor, Department of Business
Aksaray University
Address: 68100 Aksaray Turkey
E-mail: bilgehanaytac@aksaray.edu.tr

Priyanka Tripathi
Associate Professor of English, Department of Humanities and Social Sciences
Indian Institute of Technology Patna
Address: Room No. 216, 2nd Floor, Block 1, Bihta, Patna — 801103
Bihar, India
E-mail: priyankatripathi@iitp.ac.in

Apoorv Khare
Associate Professor in Marketing Management
Indian Institute of Management Nagpur
Address: VNIT Campus, South Ambazari Road, Nagpur — 440010
Maharashtra, India
E-mail: apoorv@iimnagpur.ac.in

Subhasis Ray
Professor, XLRI-Xavier School of Management
Near Circuit House Area (East)
Address: Jamshedpur — 831001, Jharkhand, India
E-mail: subhasis.ray@xlri.ac.in

In this thought leadership roundtable, the editors of the special issue, convened a critical dialogue between leading scholars in the fields of Sociology, Critical Communication Studies, Cultural Studies, Critical Management Studies, and Sustainability Studies to explore the challenges that Global South navigates in its adoption of AI. The conversation probes the power disparities between the Global North and the Global South narratives of AI and look forward to alternative forms of AI and AI Management in the Global South scenario. Highly critical and cautious of enthusiastic adaptations of AI in the Global South, the discussants prioritize the issues of Human Rights, ethics, equity, inclusivity, and resilience for the historically marginalized communities of the region.

Keywords: AI, Global South, marginal community, socio-cultural development, ethics, sustainability

Our prime concern in the Special Issue is around the nuanced socio-cultural-political impacts of AI in the Global South. This was keeping in mind the probing of a segment of society that is not only economically and structurally less privileged but even socio-culturally less represented, racially discriminated, and historically subjugated and underdeveloped. Hence, we as editors, invited critical minds researching in and with Third and Fourth world countries to offer a collaborative in-depth analysis of complex issues around the subject. The problem-solving and the strategic actionable agendas of the discussants around equitable AI development and deployment are sure to foster rigorous, responsive, and engaged scholarship. The documentation of the round table below offers an in-depth perspective, allowing for a more dynamic, immediate, and multi-voiced insight beyond traditional research articles on the issue. The experts/discussants were invited from various disciplines (sociology, communication studies, culture studies, literary studies, sustainability studies, management studies) and geographies and hence the discussion hopefully will enrich and inspire innovative research.

In this virtual roundtable, Professor Mohan J. Dutta, Dr. Bilgehan Aytac, Dr. Priyanka Tripathi, Dr. Apoorv Khare, and Professor Subhasis Ray joined the Special Issue editors (Glenn Muschert and Arindam Das) to discuss the ways AI impacts the socio-cultural political realities of the Global South. The discussion catapults the concerns around AI from its current condition in the Global South to sustainable future perspectives. The discussion also focuses on how the market actors of the Global South navigate the neoliberal impacts. The discussants probe the role of ethics in shaping a pro-Global South AI system that may instil a sense of stability and sustainability in the region. The possibilities of alternative AI discourses of the Global South that subverts and decolonizes the Global North hegemony over AI had been another vital point in the discussion. That said, and while the discussants note several cases where market subalterns' performances highlight possibilities of alternate ethics for AI from the Global South, the scepticism of the same being nothing more than purple patches persists.

AI Global Racial Capitalism and Responses from the Global South

Arindam: In an attempt to harness the positive effects of AI and yet make it equitable and just within the geopolitical, economic, and socio-cultural realities of the Global South,

we must prioritize the objectives of AI beyond "catching-up-with-the-Global-North". If the concerns of sustainability of the Global South are priority concerns, then any attempt to harness the leapfrog effects of AI must take into cognizance the same. The transformative potential of AI in the emerging economies like India should be integrated to its democratic concerns of environmental sustainability, media freedom and misinformation, religious and ethnic polarization, electoral integrity, erosion of democratic institution, protection of civil liberties, and social and economic inequality.

Mohan, you have been working for a long time on culturally-centered community-based projects of social change. What is your take on the Global South marginalized community's consumption of AI?

Mohan: I agree with you Arindam, it is important at this juncture of transformations in the Global South, including in countries such as India and Brazil, and globally that we critically interrogate the language of leapfrogging, the pursuit of extreme neoliberal development, the rise of populist authoritarianism as a mode of governmentality, and the ongoing disenfranchisement of the poor and working classes. Consider here the deeply casteist, racist, and gendered ideologies that shape algorithmic design, in global spaces of innovation such as the Silicon Valley as well as in spaces of digital development in the Global South such as Bangalore and Gurgaon. The infrastructure of caste works hand-in-hand with the infrastructures of whiteness to shape the design of digital transformations, serving the interests of global financial and knowledge capital. The layers of casteism that are reproduced in global technology organizations work alongside racial capitalism, circulating algorithms of exclusion. Moreover, the accelerated digitization of labour mediated through platforms creates spaces of extreme exploitation, with hyper-precarious workers, often from the caste margins, in sectors such as transportation and food delivery erased from spaces where they can make claims to labour rights. Consider similarly, the ongoing disenfranchisement of those at the "margins of the margins" of the Global South, with digital technologies, incorporated into the rhetoric of "smart cities" "AI hubs" etc., that are continually deployed to serve extractive agendas of a crony capitalist class, working in tandem with an authoritarian bureaucratic-political class to expel the margins from spaces of life and livelihood (Dutta, 2019). The digitization of land records is one such example of the intertwined processes of extraction and expulsion that erase the capacity of the margins of the Global South to lay claim to land. Amidst the pandemic lockdowns, the violence of erasure and expulsion experienced by migrant workers across techno-capitalist spaces of Global South digital innovation capture well the exploitative infrastructures that form the backbone of AI innovation in the Global South (Dutta, Rahman, 2022). Take the "Singapore Model" as an exemplar of the extreme neoliberalism weaved into the Global South infrastructures of digital innovations, with tactics of authoritarian surveillance and repression shaping the hyper-exploitative working conditions, poor living conditions, lack of access to fundamental necessities such as food and decent sanitation. The incorporation of digital technologies in reproducing and magnifying disinformation and hate targeting already precarious minorities within post-colonial spaces such as the targeting of Dalits, Muslims, and Sikhs in India by Hindutva,

speak to the critical role of centring justice in how we conceptualize AI, interrogating critically elite constructions of development and building spaces for voice justice (Dutta, 2019; Ramasubramanian, Dutta, 2024). Global South cities such as Bangalore, Gurgaon, or Singapore that are projected as hubs of AI innovation reproduce the extreme neoliberalism pushed by global techno-capital, further rendered potent through the workings of caste, class, and racism. Here, I am increasingly concerned about the language of de-westernization and decolonization co-opted by Global South elites to perpetuate entrenched inequalities (Dutta et al., 2021). Arindam, what do Global South AI futures possibly look like when anchored in questions of justice?

Arindam: Yes indeed! It should be our development, our modernity, and our kind of AI — one that moves beyond the restrictive discourses of Eurocentric AI that articulate a distinctly postcolonial technological ethos. Moving beyond the hegemonic narratives of the Global North AI paradigm, our AI must subversively challenge the narratives of development that focus on capitalist profiteering, extractivism, and exclusion. The habitations of the Global South modernity must generate alternate discourses of AI that prioritize well-being and inclusion. This AI must be UX-ed with purpose, intentionally designed to expand access, amplify marginalized voices, and foster deep participation by those historically excluded from technological discourses. The transformative potential of AI in the Global South lies not in mimicking the Global North but in reimagining development with justice, sustainability, and inclusion at its core. The Global-South-context-sensitive-AI-design must be adapted to the specific needs, cultures, and priorities of the Global South. The Global South AI praxis must accentuate an architecture of justice, one that interrogates and dismantles the algorithmic reproduction of casteist, racist, ageist, and gendered hierarchies. It must resist co-optation by authoritarian bureaucratic-political coalitions and crony capitalist regimes, which deploy "smart" technologies to serve extractive agendas while perpetuating the systemic disenfranchisement of the poor, Dalits, indigenous groups, and other marginalized communities. The techno-revolution of AI, intervened through the practice of care, thus can become a revolutionary project that demands centrality of justice and projects a future based on care, human dignity, and sustainability.

AI as a Leapfrogging Tool for Economic and Socio-Cultural Development in the Global South

Apoorv: AI holds great promise for global economy including the Global South. However, we must be careful about the potential pitfalls of AI while deliberating on AI as leapfrogging tool. Bias in AI systems, stemming from skewed or incomplete datasets, often reinforces societal biases related to race, gender, class, and geography, leading to systemic injustices, such as the misidentification of certain ethnic groups in facial recognition technologies. Privacy concerns are exacerbated by weak regulatory frameworks, particularly in regions vulnerable to invasive surveillance and exploitation of personal data. Furthermore, the concentration of AI development in the hands of a few global tech

corporations perpetuates neo-colonial power dynamics, sidelining the priorities of the Global South. Despite being framed as ideologically neutral, AI is influenced by the values of its creators and stakeholders, often prioritizing financial efficiency over equity and social impact. The socio-economic impact of AI is dual edged: it offers potential for democratizing access to education and healthcare but risks deepening the digital divide and exacerbating inequalities, with marginalized communities and low-income populations bearing the brunt of job losses due to automation. To harness AI equitably, context-sensitive development, diverse datasets, inclusive design practices, and strengthened local capacities are essential. Robust regulation and open public dialogue are critical to addressing ethical concerns, safeguarding data, and fostering accountability. By proactively tackling these challenges, AI can be transformed into a tool for empowerment and progress in the Global South. It is high time that we must deliberate on the critical questions such as is AI ideologically neutral and value free or a tool in the hands of global capital? What is the relation of modern tech, including AI, with socio-economic inequality?

Muschert: The notion of leapfrogging, in which regions circumvent specific developmental phases to embrace advanced technologies, presents considerable potential for the Global South. Economic and socio-cultural development can be improved by automating repetitive tasks and accelerating progress in information-driven sectors, which is facilitated by artificial intelligence. While the Global South may lag in information economy development, AI may facilitate leapfrogging stages that more developed regions underwent during previous decades, moving directly to the cutting-edge advancements of Web 3.0 and the Fourth Industrial Revolution. The prospect of leapfrogging to these new technological frontiers aligns with theories of late development, where regions that adopt new technologies later can utilize them more efficiently by learning from the experiences of earlier adopters (Comin, Hobijn, 2010). To achieve sustainable growth, these opportunities will necessitate multilateral cooperation, government support, and development a robust digital infrastructure.

On the other hand, the differential power structures in the global information economy refer back to those countries and regions with the greatest levels of technological development, especial the United States and China. Data and the value derived from its generation, processing, culling, storage, and mining remains disproportionately concentrated in areas already highly developed in information technology. Although AI offers potential for leapfrogging and development in the Global South, this could come at the risk of replicating and exacerbating existing global inequalities. As data becomes increasingly commodified in the post-industrial age, the basis of value may increasingly skew towards the countries and corporations who form the backbone of the information economy by controlling the aspects of data processing, algorithms, and platforms. Such concentration of power also extends to access to data for training large language models fundamental to generative AI systems.

Nonetheless, countertrends may also occur, often in unexpected ways. The reverse engineering of mobile money systems emerges, perhaps unexpectedly, in East Africa, allowing financial transactions via conventional mobile phone systems using SMS tech-

nology. Indeed, India is by now the global leader in digital payments, as discussed below by Dr. Apoorv.

Subhasis: Absolutely! Here, let us not forget that the Global South faces tension between rapid economic growth and severe inequality. AI can surely be of help. For example, in Brazil, AI is used for medical diagnostics in remote Amazonian regions. However, ample caution is needed to understand the implications of such leapfrogging.

Priyanka: True! AI can drive socio-cultural and economic growth in the Global South in the field of medical sciences. For example, AI can be used positively use by enhancing Assisted Reproductive Technologies (ART) access for women. It reduces ART costs, bypasses cultural barriers, and empowers women with reproductive health information. AI also offers data insights for policymakers, promoting targeted interventions. Inclusive AI applications in ART boost gender equality, fostering broader socio-economic development.

AI and Representation of Marginalized Communities in the Global South

Muschert: AI poses a significant risk to marginalized communities in the Global South, an outgrowth of the concentration of technological power in the hands of entities based in the United States and China. Linguistic minorities and smaller cultural groups are disproportionately affected by the negative stereotypes and biases that AI systems frequently propagate because of this pooling of computational power. The preeminence of English as the principal language for AI models intensifies this problem, rendering numerous communities susceptible to representation by external entities lacking comprehension of their cultural subtleties or socio-political contexts (Birhane, Prabhu, 2021).

As AI systems progressively influence public life, marginalized groups will face greater limitations in curating their own identities and representations. This digital colonialism, where data and technology from the Global North are imposed upon the Global South, undermines the agency of these communities to represent themselves authentically (Couldry, Mejias, 2019). In the Gulf Cooperation Council (GCC) nations, despite the region's financial assets, dependence on Arabic constitutes a vulnerability. Since most AI systems are optimized for English, countries relying on other languages may need help ensuring culturally relevant and accurate AI outputs. This challenge becomes accelerated for smaller language groups.

Subhasis: Our lessons from the Millenium Development Goals and the subsequent Sustainable Development Goals are clear- while there is incremental benefit to marginalized communities some of them are yet to benefit from basic skills like literacy. We must be careful about measuring and reporting of such benefits. For example, in one of our ongoing projects, we find that some members of marginalized communities can write their names and hence literate, but they cannot read their own names. We need to ensure the same phenomenon does not recur while using AI in the Global South.

Priyanka: AI's role in representing marginalized communities in the Global South, particularly concerning gender, has significant implications. Frameworks like critical race theory and intersectional feminism highlight how AI can perpetuate, or challenge inequalities based on race, class, and gender. Scholars such as Noble (2018) in *Algorithms of Oppression* argue that AI reinforces gendered and racial biases when training data lack diversity. Without representative data, AI systems risk marginalizing women, especially from minority communities, by failing to address their unique challenges. Kimberlé Crenshaw's (1989; 1991) work on intersectionality underscores how overlapping identities like gender and socioeconomic status lead to distinct experiences of marginalization. In the Global South, AI's impact on women varies with factors like caste and class, necessitating frameworks that address these complexities. Theoretical discussions (Wajcman, 2004) indicate that AI can either reinforce gender inequality or empower women, depending on its design. For example, AI tools can improve access to reproductive services, but only if tailored to the needs of marginalized women. However, AI's effectiveness is hindered by digital divides. Eubanks (2018) in *Automating Inequality* notes that socio-cultural barriers, like limited internet access, disproportionately affect marginalized women. Scholars advocate for ethical AI practices that center marginalized women's experiences. Gender-inclusive design and participatory models are vital for ensuring AI respects and uplifts their voices, aligning with Haraway's concept of Situated Knowledges (1988), which emphasizes the importance of diverse perspectives in technology development. Addressing these challenges requires designing AI systems with intersectionality in mind, using diverse datasets, and consulting marginalized groups.

Mohan: Representation in the context of AI further points to critical questions around participation, decision-making, and ownership of those at the "margins of the margins." In my work with the culture-centered approach to co-creating development solutions, I talk about the critical role of co-creating voice infrastructures in partnership with communities at the caste, race, gendered, classed margins of postcolonial development (Dutta, 2014). The ongoing processes of colonization in postcolonial India reproduce layers of extractive practices that threaten to expel Adivasis (Indigenous communities) from land and livelihood, incorporating digital technologies into the practices of development making. Simultaneously, the "margins of the margins" are incorporated into the structures of technological capital as hyper-precarious migrant workers without worker rights. In our work with low-wage migrant workers in Singapore for instance, our research team highlights the ways in which the "smart city" "eats the workers," denying them of the fundamental rights to organize, collectivize and be heard (Dutta, Rahman, 2022). Amidst the deep inequalities in the Global South that are produced by the interplays of colonialism, postcolonial development pursuing neoliberal economics, and imperialism, justice-based considerations of AI must attend to the ways in which we can co-create anti-caste, anti-capitalist, anti-imperial, working class AI infrastructures. At the level of geopolitics, it is critical to consider the ways in which Global South nations can put forth concepts of data sovereignty, infrastructural sovereignty, and economic plural-

ism in building alternative AI infrastructures that challenge US/G-7 imperialism. While engaging in this dialogue and while witnessing one of the most powerful accounts of genocide being carried out by Israel in Gaza, funded by the US, I see the critical role of building Global South AI imaginations that centre resistance to imperialism and sustain community capacities in the Global South to build just dialogues, social cohesion, and peace.

Arindam: "Justice based considerations of AI" surely must address two other major issues in the Global South — food insecurity and homelessness. An algorithmic predictive analysis may facilitate an early intervention in analysing social, economic, and demographic data in patterns of income, housing and development and thus target need before they reach crisis point. By analyzing climate and agricultural data, AI can predict crop yields and potential food shortages due to climate change or natural disasters. This information can help governments and organizations proactively address food scarcity by redirecting resources to vulnerable areas. Smart agriculture (Mendoza-Bernal et al., 2024) leads to effective traceability and management of agri-food and its supply chain (Chen et al., 2021) and thus sustainably impacting the intertwined issues of environment, economics, peace and security, and technology.

Similarly, the structural issues of homelessness in the Global South, impacted through the climate change, war/civil unrest, uninhibited globalization-urbanization, and global economic uncertainty, may be mitigated through innovative, scalable, and data-driven solutions. Real-time AI models, predictive analysis and ML can provide timely data towards preventive measures in mapping and addressing homelessness. Researchers at IIT Kharagpur (Raj and Mitra, 2024) used satellite imagery (Very-High-Resolution [VHR] Imagery) and AI algorithms (deep learning) to map slum areas in Indian cities. The project focused on identifying unplanned urban settlements, which often house homeless or semi-homeless populations, enabling targeted policy interventions.

Strategies for Promoting AI Inclusion and Addressing Digital Inequality in the Global South

Muschert: Yet, as the same time as AI becomes more widespread, the digital divide may widen, as it has historically been defined by unequal access to the internet and digital technologies (Hilbert, 2011). The exclusion of vulnerable populations from AI, such as women, the elderly, rural residents, and linguistic minorities, has the potential to exacerbate existing social and economic disparities. Education and skill development are essential in addressing this divide. Programs aimed at equipping subaltern populations with AI literacy can mitigate exclusion, and mobile technology could serve as a conduit for AI access in regions where mobile phone usage is widespread but access to personal computers is limited. Given the ubiquity of smartphones, AI functionalities could be integrated into mobile platforms to democratize access (Donner et al., 2011). Nevertheless, the establishment of the requisite infrastructure for equitable access will necessitate the

collaborative efforts of governments, NGOs, and the private sector to achieve widespread AI inclusion.

Subhasis: Strategies for using AI to reduce inequality need to consider the principles of justice and fairness. Deployment strategies need to consider local political and institutional structures in a way that AI can bypass existing bottlenecks for technology adoption. One may consider community driven AI initiatives and use appropriate structures for the same. Local co-operatives, for example, can play a role, in involving marginalized farmers and provide solutions for climate resilient agriculture.

Priyanka: Promoting AI inclusion and addressing digital inequality for women in the Global South involves increasing access to affordable internet, enhancing digital literacy, and designing AI tools with cultural sensitivity. Empirical data highlights that only 35% of women in low-income countries have internet access, emphasizing the need for targeted policies (ITU, 2021). Additionally, community-based digital skills programs have proven effective in reducing the gender digital divide (UN Women, 2020). Telehealth is also essential, offering accessible healthcare in remote areas and reducing travel barriers (WHO, 2021).

Arindam: Yet, one must be cautioned that AI-inclusivity and integration can inadvertently lead to surveillance and hegemony if not carefully implemented and regulated. With marginal and underrepresented community's lesser control and access over their own AI-generated data, surveillance may be expanded in sites that previously had autonomy over their privacy. Moreover, AI systems, even when inclusive, are often developed using frameworks and values rooted in dominant global or regional cultures. This can subtly enforce the norms, languages, and ideologies of hegemonic powers under the guise of inclusivity.

Mohan: The question of inequality and AI must begin by recognizing that AI, in its organizing structure, is a feature of global racial capitalism (Dutta, 2021). Global technology capital forms the frontiers of the extractive and exploitative industries, continually at work to reproduce inequalities and to legitimize these inequalities on a global scale. As an extension of global technology capital, AI creates new frontiers for capture and colonization, building architectures of surveillance, commoditization of participation, and repression of resistance. It is worth noting here the ways in which the language of digital divide and building technological access forms the organizing logic of global technology capital, creating new spaces for capture, exploitation and oppression. Under the guise of building free access, global technology corporations create new zones of surveillance and marketization of human participation. In the Global South, the performance of developing skills, creating technological empowerment, and building accessibility is integral to the creation of new zones of profiteering. Ironically then, global technology capital reconfigures the language of technology, empowerment and participation to serve the interests of capital. For example, under the guise of sustainability, the positioning of net zero creates new zones of technology-based extraction, rendering invisible the processes of extraction and exploitation that form the infrastructures of the technologies.

Implications of AI Adoption for Socio-Cultural Dynamics in the Global South

Muschert: Adopting AI in the Global South will likely transform socio-cultural dynamics. On the one hand, AI could accelerate the modernization of economies and create new opportunities for social mobility, as seen in sectors like digital banking and agriculture (Mayer-Schönberger, Cukier, 2013). On the other hand, AI threatens to homogenize cultural identities by reinforcing dominant cultural norms, particularly those from the Global North. The threat of cultural homogenization arises from the design and implementation of AI systems, which frequently reflect the developers' biases. AI's reliance on Western-centric datasets and algorithmic logic could marginalize non-Western thinking and cultural practices (Birhane, 2021). Additionally, the proliferation of AI systems may lead to a decrease in linguistic diversity, as it becomes increasingly difficult for smaller language groups to safeguard their cultural heritage in the presence of technologies that are intended for dominant languages.

Bilgehan: Turkish context is always interesting, considering it has long been a candidate of the European Union and gets secular day by day but doesn't refuse its imperial and religious historical background. It has been governed by a religious party for 20 years. When talking about the Global South and Global North, one can say the country stays in limbo. However, there are many more reasons for us to position Turkey in the Global South. The entire lands of the country were occupied by colonial powers and fought a war of liberation only 100 years ago, and today, we see those Western capitalist values that culturally dominate social and cultural life just as they do other previous colonies. In Turkey, whenever a "next big thing" starts to be discussed, whether in academia, media, or daily chats, we see proactive efforts (we can even call them futuristic approaches) remain very individualistic. This mostly stems from the learned helplessness that comes from being from an 'underdeveloped' country. For example, in Turkey, as early as 1950, we see a Turkish scientist, Cahit Arf, give a lecture in a city called Erzurum, which is not a metropolitan capital, and its main title was "Can Machines Think and How Can They Think?". We see that such people are very unlikely to take the lead in the long term in the global south. I want to frame this within *historical trauma on learned helplessness* (Chery, 2021). I am not referring to conspiracy theories, but rather to the broader societal reactions and tendencies. The Turkish people may now pay tribute to Cahit Arf and print his pictures on banknotes today, but it would not be unfair to say that they have never been inclined to follow idealistic people with such futuristic perspectives (Sarı, 2021). There is a more obvious example in Turkish history: the first national car project and strong representative of the national capital against global capital, Devrim (Revolution). The pessimistic attitudes within Turkish society and how the project's engineers managed these sentiments were widely recognized, even becoming the subject of a notable film (Dönmez, Keşaplı, 2021). This fact was well described very early by Turkish sociologist Cemil Meriç poetically as follows*:*

> *(...) His European friends looked at the old man with pity and whispered in his ear: 'No, young man,' they whispered, 'you are underdeveloped.' And our intellectuals*

> *proudly embraced this death penalty label attached to our chests by the Christian West as if it were a 'legion of honor'* (Meriç, 1985).

So, can we talk about the same self-doubt and learned helplessness in the attitude toward artificial intelligence? As it turns out, the answer to this question is 'yes'. Recently, one of my colleagues at my institution measured the artificial intelligence anxiety of management information systems students. The findings were quite unexpected, given the students' extensive familiarity with technology. The results revealed that they are seriously anxious about AI. Their anxiety was mostly related to their competency and losing their job in the future (Aytaç, 2022). To be honest, I sometimes share similar feelings. While it is difficult to claim that our concerns encompass issues such as surveillance and digital colonialism emerging from big data and AI technologies, perhaps they should. Maybe we can uncover these more nuanced concerns through qualitative inquiries; however, at present, I do not observe significant awareness of these issues within large parts of society, the media, and academia.

Apoorv: The entry by Bilgehan made me think about an important and related issue which has wider implications but is particularly relevant for Global South.

India is leading the world in digital transactions. Its Unified Payment Interface (UPI) is the world's leading digital payment platform (*Times of India*, 2024). A relatively less discussed but widely felt effect of the growth in digital transactions is increasing digital, specifically UPI related frauds (*Business Today*, 2024). This is posing new challenges, especially rendering a large population in India vulnerable. This vulnerability in a fast-digitising marketplace is no longer limited to primarily less educated, migrant, subaltern population. The potential vulnerability is now extended to a large population with digital footprint. Many, who in a non-digital world would not be considered vulnerable, have been experiencing failed or sometimes successful attempts by digital imposters frequently. Despite various attempts by Government agencies, digital fraud and digital vulnerability, are increasing rapidly. Digital vulnerability is defined as "a universal state of defencelessness and susceptibility to (the exploitation of) power imbalances that are the result of increasing automation of commerce, datafied consumer–seller relations, and the very architecture of digital marketplaces" (Helberger et al., 2022: 176).

The new digital technologies and data analytics can create new power imbalances. "Persuasion profiles", "Dark Patterns" are examples of design choices with potential to create new vulnerabilities (Helberger et al., 2022). Weak policy and institutional infrastructure, a weaker enforcement, and a general lack of user- or citizen- centric environment makes the situation grim for Global South.

Priyanka: The intersection of artificial intelligence (AI) and socio-cultural contexts is a dynamic, evolving field of study. As AI technologies advance, examining their impact on diverse societies — especially in the Global South — is crucial. This region is not only economically and structurally less privileged but also socio-culturally underrepresented, racially discriminated, and historically marginalized. AI development is predominantly concentrated in the techno-progressive Global North, with most research focus-

ing on resource-rich Euro-American communities. The World Economic Forum (2023) has highlighted this "AI divide" between the Global North and South. While the Global South seeks to leverage AI in agriculture, healthcare, education, climate action, poverty alleviation, and economic growth, it faces challenges related to data infrastructure, governance, AI usability, and accessibility. Additionally, understanding how cultural perspectives shape AI's ethical guidelines, perpetuate biases, influence cultural production, and affect economies is essential. Finally, it's critical to explore how AI-driven neoliberal colonialism and capitalism are resisted and negotiated by Global South communities. For instance, AI adoption in the Global South offers potential benefits for doubly marginalized women by improving reproductive autonomy through tailored telehealth services. However, challenges remain. Theoretical frameworks like intersectionality reveal that AI might reinforce biases if not carefully implemented, risking further marginalization (Noble, 2018). Empirical studies also show digital divides persist, limiting access to AI tools (GSMA, 2020). Thus, while AI can enhance reproductive rights by providing private, accessible healthcare information, equitable access and cultural sensitivity are essential for meaningful impact (UN Women, 2020).

Importance of Cultural Diversity and Values in the Global South in shaping the Ethical Standards of AI

Mohan: Global South offers a register for centring justice when we foreground questions of worker rights, Indigenous rights, gender rights, and minority rights in how we develop ethical frameworks around AI. In my own work with the culture-centered approach (CCA), everyday friendships with communities at the "margins of the margins" serve as registers for conceptualizing ethics and technology, anchoring conversations around ethics in community voices and in struggles for social justice (Dutta, 2021). The voices of the Global South, in laying claims to ownership of knowledge, sovereignty of stories, ownership of data, and actually existing socialist politics, create diverse registers for how we conceptualize AI ethics. Cultural diversity and diverse values thus return to questions around building a new world information communication and technology infrastructure, continuing with the earlier work of the non-aligned movement (NAM) around the new information communication order. The processes of participation of communities at the margins, created through partnerships in solidarity, de-center the hegemonic registers of ethics that are emergent from the Global North.

Muschert: An ethical challenge in deploying AI is the risk of eroding cultural diversity. The reliance on computational logic and English-centric computing asserts an environment where local values and perspectives become sublimated. In many cases, this may result in a homogenization of cultures, where the values of dominant technological actors overshadow the cultural richness of smaller communities (Couldry, Mejias, 2019). Maintaining cultural diversity is crucial for the preservation of local traditions and for fostering innovation. Diverse perspectives are often the source of new ideas and ethical frameworks to challenge dominant narratives. As such, ethical standards for AI should

reflect cultural contexts, thereby incorporating values from various cultural perspectives, ensuring that no single perspective dominates the ethical discourse surrounding AI (Floridi, Cowls, 2019).

Priyanka: In India, the ethical standards of AI must consider cultural diversity, as illustrated by Amartya Sen's (1999) capability approach, which prioritizes individual freedoms and socio-cultural contexts. This framework highlights that ethical AI should reflect local values, ensuring it addresses specific societal needs, such as gender equality and social justice, rather than imposing a uniform, Western-centric model that risks exacerbating existing inequalities. With nearly 70% of the population in rural areas, AI must address urban-rural divides in access and literacy to avoid fear and mistrust.

Arindam: I agree with Priyanka and think that we must acknowledge the varied habitations of AI ethicalities. While the Global North has historically dominated the discourse on AI ethics, the Global South surely has its own critical perspectives to offer that challenge and enrich this conversation. Moving beyond the individuated privacy-centric White Western worldview, the AI of the Global South may find a better scope in the community centricity. This community centricity will help in making the AI infrastructure more inclusive, equitable, and empathetic. Moreover, acknowledgement of epistemic pluralities in AI ethics, beyond western hegemonic standards of knowledge, will lead to the process of decolonizing the ethical imperialism in AI by the Global North. The demands of justice and equity by the Global South runs counter to the 'just' hegemonic access and control of big data by the capitalist Global North digitalities. Ethical standards shaped by the values of the Global South would prioritize equitable data governance, ensuring that local communities retain control over their data and benefit from its use/representation. Further, the ethical knowledge systems (essentially mediated through the ethos of sustainability) of various indigenous communities of the Global South may go to positively inform the AI infrastructure. Indeed, researchers have called for decolonizing the western model of AI (Adams 2021; 2025) and promulgating an alternate AI from within the frameworks of the indigenous knowledge systems (Lewis et al., 2024).

Impact of Socio-Cultural, Demographic, and Economic Factors on AI Accessibility and Usability in the Global South

Muschert: A range of socio-cultural, demographic, and economic factors shape the accessibility and usability of AI technologies in the Global South. The rapid advancement of AI requires a workforce capable of adapting to the needs of an AI-driven economy. However, there often may be a mismatch between the skills available in the Global South and the demands of the global AI industry. Many workers may need more technical education to thrive in AI-related industries, creating a skills gap that limits the update and utility of AI in the Global South (Hilbert, 2011).

Cultural factors may also play a role. In some communities, there may be skepticism toward AI technologies due to religious or philosophical beliefs. For example,

studies show that religiosity can influence attitudes toward AI, with more religious individuals often expressing increased anxiety about AI technologies (Bryson, Theodorou, 2019). Addressing such sociocultural barriers requires targeted educational programs and community engagement initiatives that respect local beliefs while promoting digital literacy.

Mohan: The hegemonic approach to digital technologies, including the approach to AI, constructs communities at the margins in deficit, framing the questions of access and usability around barriers. Culture is treated as an essence, and efforts of digital literacy and design are then organized around making these technologies culturally sensitive (Dutta, 2007). I have argued that such culturally sensitive approaches to technology design reproduce the colonial-imperial-racial capitalist structures by incorporating the participation of those at the margins into the ambits of global technology capital. The rhetoric of "uplifting the burden" continues to be reproduced while actually participating in the hyper-exploitation of the precarious classes. Concepts of co-design, participation, and community engagement are incorporated to serve the profiteering agendas of global techno-capital. We need to critically engage our approaches to empowerment and education around AI, examining the underlying ideologies that take for granted AI futures, devoid of questions of justice, worker rights, minority rights, indigenous rights, transgender rights, women's rights, etc. Our activist and community partnerships at the Center for Culture-centered Approach to Research and Evaluation (CARE) offer a culture-centered approach to AI, asking critical questions such as, how do we sustain the power of communities at the margins to retain communication sovereignty? How do we sustain the power of communities to participate in the design and development of AI technologies, anchored in their voices? What would just AI look like, when embedded within struggles against imperialism, settler colonialism, patriarchy, and racial capitalism?

Bilgehan: I would like to address the issue again from a different point of view in the context of my country. Religiosity has always been at the heart of socio-political debates in Turkey. Although there has been a political power in the country for 20 years that has adopted the discourse of religious idealism, we are faced with an increasingly secularized youth. Still, religious life is quite determinant within the daily life of society. So, my colleagues and I decided to study how religiosity affects attitudes towards artificial intelligence (Aytaç et al., 2024). I think no one in an Islamic society has initiated something similar to this before us. Most of the participants were students or white-collar workers. In other words, the sample's education level wasn't low at all. Our findings showed us that religiosity positively predicted AI anxiety, and this relationship was mediated by populism. In parallel with this, religiosity negatively predicted intention to use ChatGPT. I think the results are quite alarming. We see that populist political discourse and populist attitudes towards science have become widespread all over the world, most naturally in the Global South, and this discourse is heavily fed by religiosity. This study demonstrates that there is a populist attitude toward artificial intelligence that aligns closely with the prevailing attitudes toward science.

The Role of "Ethics of Care" in AI Deployment in the Global South

Bilgehan: Another significant issue in AI ethics is labor exploitation. As companies increasingly invest in AI technologies, there are alarming instances, such as in Venezuela, where data-labeling firms have taken advantage of the economic crisis to hire workers at extremely low wages (Hao, Hernández, 2022). This exploitation not only raises ethical concerns but also highlights the need for stronger protections and fair compensation for workers in the AI sector.

Mohan: AI, as the frontier of digital innovations, is very much a critical resource in the global racial capitalist, imperial, settler colonial infrastructure. In this sense, the very design of AI exists in opposition to the ethic of care anchored in justice. In our work with hyper-precarious platform workers across diverse global contexts, our research team at the Center for Culture-centered Approach to Research and Evaluation (CARE, Massey University, New Zealand), documents the ways in which automated platforms continually work to marginalize workers, creating communicative inequalities and disenfranchising workers from spaces of collectivization. Automated forms of work management around transportation and food delivery while on one hand produce extractive models of profiteering for global capital, on the other hand, they erase workers from spaces where they could raise claims to workers' rights to care and decent pay. Our research noted that amidst the pandemic, digitally automated systems of interaction and workplace management heightened the risks borne by the workers, without access to adequate protections such as face masks and sanitizers. Juxtapose this in the backdrop of the mutual aid and relationships of care that are co-created by workers in challenging the alienating digital structures. An "ethic of care" anchored in the Global South turns toward alternative imaginations that challenge the capitalist consolidation of innovations such as AI. Worker run platforms, platform cooperatives, worker organizing in the Global South (particularly Global South technology worker organizing) offer some resistance of hope against the hegemonic imaginaries of the smart digitalities (Dutta, Kaur-Gill, 2018).

Muschert: The ethics of care, which emphasizes relationality, interdependence, and the well-being of vulnerable populations, provides a compelling framework for the ethical deployment of AI in the Global South. This approach contrasts with the more utilitarian ethical frameworks often used in AI development, which prioritize efficiency and productivity over the well-being of individuals (Held, 2006). In the context of the Global South, where economic inequalities and power imbalances abound, an ethic of care approach could ensure that AI technologies prioritize marginalized communities' needs. By design, AI systems used in healthcare or social services should enhance, rather than replace, human relationships. By focusing on the well-being of individuals and communities, an ethics of care can guide the development of AI systems that are more equitable and responsive to the needs of the Global South (Tronto, 1993).

Priyanka: The "ethics of care," grounded in relational and feminist theories, focuses on empathy, community welfare, and contextual awareness (Gilligan, 1982). In the Global South, this framework can steer AI deployment by emphasizing local values and needs,

thus enhancing inclusivity. However, if not implemented thoughtfully, AI risks worsening existing inequalities. To counter stratification, it is essential for stakeholders to involve local communities in AI development, ensuring that ethical frameworks are relevant to their contexts and effectively address power imbalances. This promotes equitable access to technology and resources, fostering social cohesion. The ethics of care is seen as both a practice and a principle, understood as a "motive, ideal, virtue, and method" (Sander-Staudt, Hamington, 2011). Care theorists agree that this approach requires a relational morality that responds to specific individuals beyond rigid rules or outcomes (Hamington, 2019). By considering context — such as space, time, and culture — those working with AI can better understand the consequences of their actions on others, thereby reducing moral distance. Developers and deployers of AI must ensure algorithms consider relevant variables, such as a person's historical and cultural background. For instance, in learning analytics, context can significantly influence a student's GPA, particularly in under-resourced areas. Additionally, in the case of the Amazon delivery bot, former managers indicated that "the largely automated system is insufficiently attuned to the real-world challenges drivers face every day" (Soper, 2021). Understanding vulnerability is crucial within the ethics of care, as it underscores the importance of recognizing the needs and suffering of others. AI development based on this principle should ensure that algorithms do not obstruct individuals from fulfilling their needs. This entails making sure that algorithms support the requirements of protected groups and marginalized stakeholders, avoiding the exploitation of vulnerabilities as data points. For example, admissions processes should consider the specific challenges faced by students with attention deficit disorders, acknowledging that their capabilities may vary across subjects.

AI and the Sustainability of Communities in the Global South

Bilgehan: At a recent international conference, I attended, an Iranian student openly expressed these concerns, saying "you have not lived in a dictatorship, and you don't know what it is like, I have, and the dictatorship of artificial intelligence will be the most challenging of all". Here, we anticipate that the fear created by the anti-democratic political life, which is more common in the global south, will be multiplied by artificial intelligence, and this may draw our discussion into a completely different context. It is well known that mass surveillance tools in China intensified such concerns during the COVID-19 pandemic (Yu, 2020). According to the Global Surveillance Index (GSI), at least 75 of the 176 countries worldwide are investing in and using artificial intelligence for surveillance, particularly in areas like smart cities, facial recognition, and smart policing. We should interpret this not just in a local but in a global context. Artificial intelligence technologies can increase not only local autocratization but also the surveillance and control powers of global powers over these countries. According to the same report, Chinese product pitches often come with soft loans to incentivize governments to buy their equipment, particularly in countries like Kenya, Laos, Mongolia, Uganda, and Uzbekistan, where access to such technology would be limited. This raises concerns

about China's role in subsidizing advanced repressive technologies (Feldstein, 2019). Further, locally, it is expected that these systems will reinforce military surveillance systems, which can lead to *discriminatory practices and unaccountable military activity on a political and legal level* (Saheb, 2023: 373). Obviously, for people in the Global South, this poses an even greater risk. We can find support for this just by looking at this statistic: Since 1950, Africa has seen the most military coups globally, with 214 out of 486 attempts, at least 106 of which were successful (AJLabs, n.d.). Nonetheless, this situation is troubling not only for the cultural, diplomatic, and economic dominance of the Global North over the Global South but also for the nations within the Global North itself. We are witnessing a global *democratic recession,* and this is reflected in various numerical indicators in countries such as Germany, Belgium, Finland, and France. AI surveillance systems have the potential to boost this deterioration. In particular, this recession is said to have accelerated in the post-pandemic period (Balmori de la Miyar, 2021).

Mohan: The concept of sustainability of communities in the Global South is deeply intertwined with justice. Yet, questions of justice are often erased from liberal elite spaces that define sustainability in the Global South and hold power and control over spaces of sustainability. The accelerated proliferation of neoliberal policies has worked alongside the deployment of technology-enabled solutions (such as AI), framed as the future of postcolonial development. Yet, consider the large-scale corporate take-over of Indian agriculture in the pursuit of neoliberal policies that have erased farmers, large proportion of whom are women, from their spaces of livelihood. The pursuit of neoliberal policies in agriculture, framed under the guise of sustainability, has shaped the ongoing expulsion of small-scale farmers from their agricultural practices, creating conditions of food insecurity, and shaping the pandemic of farmer suicides. As witnessed in the work of our community partner Deccan Development Society (DDS), organized in the form of women's cooperatives (*sanghams*), the ownership of communities over storytelling resources (such as community radio) is a critical element in securing seed sovereignty, and in securing the sustainability of communities in growing healthy food. These culturally-centered struggles for sustainability in community life are organized in opposition to hegemonic agendas (such as net zero, carbon offsetting, carbon credits etc.) that co-opt sustainability to serve the agendas of postcolonial elites and global technology capital.

Muschert: The sustainability of communities in the Global South is a critical issue, and AI has the potential to both enhance and undermine community sustainability. On the positive side, AI can optimize resource management, improve healthcare outcomes, and increase access to education, all of which contribute to the long-term sustainability of communities (Mayer-Schönberger, Cukier, 2013). However, there are also significant risks. AI-driven surveillance technologies, for example, might suppress legitimate dissent and facilitate the control of populations, raising concerns about the role of AI in perpetuating power imbalances (Kwet, 2019). Additionally, the economic benefits of AI may be distributed unevenly, with wealthier, more technologically advanced regions benefiting disproportionately. AI deployment must follow policies that promote equitable access to its benefits and protect human rights, thereby ensuring sustainability in the Global South.

Ray: Deployment of AI in the Global South needs to be cognizant of the ground reality. Emerging economies run on large unorganized labor. AI usage can create labor market disruptions and affect the livelihood of millions of frontline workers involved in routine jobs. Instead, we need to work on areas like healthcare where AI can have fast and positive impact. AI needs to be driven by national developmental priorities rather than corporate interests. Nations in the global south need to identify priority sectors for AI solution development with a clear agenda to address embedded inequalities.

Apoorv: AI holds great potential to alleviate many problems humanity is facing, especially in the Global South in the areas of welfare, healthcare, agriculture, economy etc. However, Michael Foucault's famous statement, "not that everything is bad, but that everything is dangerous" succinctly reminds us of the dangers of AI if it is left unchecked. We highlight the need to critically engage with the questions of colonization, socio-economic inequalities, gender bias, representation and sustainability of marginalized communities in the age of AI while appreciating the promise of AI.

Priyanka: AI has the potential to greatly enhance community sustainability in India by streamlining resource management, advancing agricultural practices, and boosting local economies. Theoretical frameworks such as Actor-Network Theory (Latour, 2005) emphasize the interconnectedness of technology and social systems. Data shows that AI applications in agriculture can boost yields by as much as 30% in India (NITI Aayog, 2021). However, ensuring equitable access and engaging communities are vital to addressing local needs and preventing increased inequalities. While developed nations utilize AI to create algorithms that foster economic growth, the Global South is experiencing a rise in industries that depend on low-skilled labor for data labeling and correction in the AI value chain. To help these nations escape this pattern, local governments and international aid organizations must evaluate current capabilities and strategically plan investments to enhance their development trajectory and deliver sustainable benefits. For instance, investments should target improving essential AI infrastructure and skills to lay the groundwork for effective solutions.

Arindam: No wonder this thought leadership roundtable culminates around the discourse of sustainability and Global South AI. We all have strongly endorsed the idea that AI from Global South must decolonize the digital discourses of the Global North and project towards a more sustainable future. By situating AI within the broader frameworks of equity, ethics, and decolonization, the discussions illuminate the urgent need for alternative, context-sensitive approaches that challenge hegemonic Global North narratives. The insights gained emphasize the importance of fostering AI systems that prioritize stability, inclusion, and resilience, particularly for communities historically marginalized by structural inequities.

The dialogue in which we engage strongly underscores the potential of AI both to address immediate socio-economic disparities and to lay the groundwork for long-term sustainable development in the Global South. Yet, we extend great cautiousness for accepting any overly optimistic interpretations and ever urge a critical engagement with market subaltern performances and the risk of superficial solutions that fail to address systemic issues.

References

Adams R. (2021) Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*. Vol. 46, no 1-2, pp. 176-197.

Adams R. (2025) *New Empire of AI: The future of global inequality,* Oxford: Polity Press.

AJLabs (n.d.) (2024) Mapping Africa's coups d'etat across the years. *Al Jazeera*. Retrieved September 25, from https://www.aljazeera.com/news/2023/8/30/mapping-africas-coups-detat-across-the-years

Aytaç Z. (2022) Üniversite öğrencilerinin yapay zekâ öğrenme ve iş değiştirme kaygılarının otonom araçlar ve akıllı evler özelinde değerlendirilmesi. *Üçüncü Sektör Sosyal Ekonomi Dergisi,* vol. 57, no 4, pp. 2975-2989.

Aytaç Z., Aytaç M. B., Yıldırım E. (2024) *GOD SAVE US FROM AI! Religiosity and intention to subscribe to ChatGPT*. Manuscript submitted for publication.

Balmori de la Miyar J. R. (2021) Are OECD countries in a rule of law recession? *Law and Development Review,* vol. 14, no 2, pp. 401-428.

Birhane A., Prabhu V. U. (2021) Algorithmic injustice: A relational ethics approach. *Patterns*, vol. 2, no 2, p. 100205.

Bryson J. J., Theodorou A. (2019) How society can maintain human-centric artificial intelligence. *Human-centered digitalization and services* (Marja Toivonen and Eveliina Saari eds.), Singapore: Springer Translational, pp. 305–323.

Business Today (2024) https://www.businesstoday.in/technology/news/story/digital-payment-frauds-surge-in-india-as-upi-transactions-skyrocket-rbi-report-431695-2024-06-01 (Accessed on 01/10/2024).

Chen H., Chen Z., Lin F., Zhuang P. (2021) Effective management for blockchain-based agri-food supply chains using Deep Reinforcement Learning. *IEEE Access*, vol. 9, pp. 36008–36018.

Chery L. (2021) *The effects of historical trauma on learned helplessness and self-fulfilling prophecy: A conceptual paper for UNESCO and the United Nations Community*. UNESCO Chair in Community, Leadership, and Youth Development, Penn State University. https://agsci.psu.edu/unesco/publications/white-papers/the-effects-of-historical-trauma-on-learned-helplessness-and-self-fulfilling-prophecy/the-effects-of-historical-trauma-on-learned-helplessness-and-self-fulfilling-prophecy-chery.pdf

Comin D., Hobijn B. (2010) An exploration of technology diffusion. *American Economic Review*, vol. 100, no 5, pp. 2031-2059.

Couldry N., Mejias U. A. (2019) *The costs of connection: How data is colonizing human life and appropriating it for capitalism,* Redwood City, CA: Stanford University Press.

Crenshaw K. (1989) Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, no 1, pp. 139-167.

Crenshaw Kimberle (1991) "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color." *Stanford Law Review,* vol. 43, no 6, pp. 1241–1299.

Dönmez A., Keşaplı O. (2021) Devrim Arabaları (2008) ve Usta (2009) filmlerinde küresel sermaye eleştirisi ve ulusal sermaye temsili. *Türkiye Medya Akademisi Dergisi,* vol. 1, no 1, pp. 83-98.

Donner J., Gitau S., Marsden G. (2011) Exploring mobile-only Internet use: Results of a training study in urban South Africa. *International Journal of Communication*, no 5, pp. 574–597.

Dutta M. J. (2014) A Culture-Centered Approach to Listening: Voices of Social Change. *International Journal of Listening*, vol. 28, no 2, pp. 67–81.

Dutta M. J. (2019) Digital transformations, smart cities, and displacements: Tracing the margins of digital development. *International Journal of Media Studies*, vol. 1, no 1, pp. 1-21.

Dutta M. J., Rahman Md. M. (2022) The city eats the worker: migrant negotiations of COVID-19 and resistance amidst the COVID-19 crisis. *Consumption Markets & Culture*, vol. 26, no 3, pp. 194–209.

Dutta M. J. (2021) "Universities, civility, and expression in the age of new media." *Civility, Free Speech, and Academic Freedom in Higher Education: Faculty on the Margins* (R. Ballerstadt, Bhattacharya eds.), New York: Routledge.

Dutta M., Ramasubramanian S., Barrett M., Elers C., Sarwatay D., Raghunath P., ... & Zapata D. (2021) Decolonizing open science: Southern interventions. *Journal of communication*, vol. 71, no 5, pp. 803-826.

Eubanks V. (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor,* St. Martin's Press.

Feldestein S. (2019) "The global expansion of AI surveillance." Carnegie Endowment for International Peace. https://carnegie-production-assets.s3.amazonaws.com/static/files/files__WP-Feldstein-AISurveillance_final1.pdf.

Floridi L., Cowls J. (2019) A unified framework of five principles for AI in society. *Harvard Data Science Review*, vol. 1, no 1.

Gilligan C. (1982) *In a Different Voice: Psychological Theory and Women's Development,* Harvard University Press.

GSMA (2020) *The Mobile Gender Gap Report 2020*. Retrieved from GSMA.

Hamington M. (2019) Care Ethics and the Politics of Care. *The Cambridge Handbook of the Ethics of AI*.

Hao K., Hernández A. P. (2022, April 20) The AI industry runs on the invisible labor of humans. *MIT Technology Review*. https://www.technologyreview.com/2022/04/20/1050392/ai-industry-appen-scale-data-labels/

Haraway D. (1988) Situated Knowledges: The Science Studies Reader. *Feminist Studies*, vol. 14, no 3, pp. 575-599.

Helberger N., Sax M., Strycharz J. et al. (2022) Choice Architectures in the Digital Economy: Towards a New Understanding of Digital Vulnerability. *J Consum Policy*, no 45, pp. 175–200.

Held V. (2006) *The ethics of care: Personal, political, and global,* Oxford, UK: Oxford University Press.

Hilbert M. (2011) Digital divide 2.0: "The great leap" toward universal access to digital technologies. *Telecommunications Policy*, vol. 35, no 8, pp. 601–620.

International Telecommunication Union (ITU) (2021) Measuring digital development: Facts and figures 2021. Retrieved from ITU.

Kwet M. (2019) Digital colonialism: US empire and the new imperialism in the Global South. *Race & Class*, vol. 60, no 4, pp. 3–26.

Latour B. (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory,* Oxford University Press.

Lewis J. E., Whaanga H., Yolgörmez C. (2024) Abundant intelligences: placing AI within Indigenous knowledge frameworks. *AI & Sociaty*. https://doi.org/10.1007/s00146-024-02099-4

Mayer-Schönberger V., Cukier K. (2013) *Big data: A revolution that will transform how we live, work, and think,* Boston, MA: Houghton Mifflin Harcourt.

Mendoza-Bernal J., González-Vidal A., Skarmeta A. F. (2024) A convolutional neural network approach for image-based anomaly detection in Smart Agriculture. *Expert Systems with Applications*, no 247, p. 123210.

Meriç C. (1985) *Bu ülke,* İletişim Yayınları.

Mohan J. Dutta, Satveer Kaur-Gill (2018) "Mobilities, Communication, and Asia| Precarities of Migrant Work in Singapore: Precarities of Migrant Work in Singapore: Migration, (Im)mobility, and Neoliberal Governmentality." *International Journal of Communication,* no 12, pp. 4066-4084.

Mohan J. Dutta (2007) Communicating About Culture and Health: Theorizing Culture-Centered and Cultural Sensitivity Approaches. *Communication Theory*, vol. 17, no 3, pp. 304–328.

NITI Aayog. (2021) *National Strategy for Artificial Intelligence*. Government of India. Retrieved from NITI Aayog.

Noble S. U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism,* NYU Press.

Raj A., Mitra A. (2024) *Deep Learning for Slum Mapping in Remote Sensing Images: Meta-analysis and Review* https://arxiv.org/html/2406.08031v1.

Ramasubramanian S., Dutta M. J. (2023) The CODE^SHIFT model: a data justice framework for collective impact and social transformation. *Human Communication Research*, vol. 50, no 2, pp. 173-183.

Saheb T. (2023) Ethically contentious aspects of artificial intelligence surveillance: A social science perspective. *AI and Ethics,* vol. 3, no 2, pp. 369–379.

Sander-Staudt M., Hamington M. (2011) The Ethics of Care: A Feminist Perspective. *Feminist Ethics and Social Theory*.

Sarı F. (2021) Cahit Arf'in "Makine düşünebilir mi ve nasıl düşünebilir?" adlı makalesi üzerine bir çalışma. *TRT Akademi,* vol. 6, no 13, pp. 812-833.

Sen A. (1999) *Development as Freedom,* Oxford University Press.

Soper S. (2021) Amazon's delivery bot is a 'lack of compassion,' ex-managers say. *Business Insider*.

Times of India (2024) https://timesofindia.indiatimes.com/technology/tech-news/upi-surpasses-worlds-leading-digital-payments-platforms-with-this-record/articleshow/112949499.cms (Accessed on 01/10/2024)

Tronto J. C. (1993) *Moral boundaries: A political argument for an ethic of care,* Abingdon-on-Thames, UK: Routledge.
UN Women (2020) The Impact of COVID-19 on Women and Girls: The Shadow Pandemic. Retrieved from UN Women.
Wajcman J. (2004) *Technofeminism,* Oxford: Polity Press.
World Economic Forum (2023) *The AI Divide: Bridging the Global North-South Gap*. Retrieved from World Economic Forum.
World Health Organization (WHO) (2021) Telehealth: A new way of providing care during the COVID-19 pandemic. Retrieved from WHO.
Yu A. (2020) Digital surveillance in post-coronavirus China: A feminist view on the price we pay. *Gender, Work & Organization,* vol. 27, no 5, pp. 774-777.

Влияние, проблемы и перспективы искусственного интеллекта на глобальном Юге: свежие идеи и передовые практики. Круглый стол

Ариндам Дас
Профессор Школы свободных искусств Университета Аллайанс
Адрес: Чиккахагаде-Кросс, главная дорога Чандапура-Анекал
Анекал, Бангалор — 562106, Карнатака, Индия
E-mail: arindam.das@alliance.edu.in

Гленн Масхерт
Профессор Факультета общественного здравоохранения и эпидемиологии в Колледже лечебного дела и медицинских наук Университета науки и технологий в Халифе
Адрес: Почтовый ящик 127788 Абу-Даби, ОАЭ
E-mail: glenn.muschert@ku.ac.ae

Мохан Дж. Датта
Декан и профессор Факультета коммуникации в Школе коммуникаций, журналистики и маркетинга Университета Мэсси и директор Центра культурно-ориентированного подхода к исследованиям и оценке (CARE)
Адрес: До востреб. 11 222, Палмерстон-Норт, 4442 Новая Зеландия
E-mail: mohanjdutt@gmail.com

Мухаммед Бильгехан Айтеч
Доцент Факультета бизнеса в Университете Аксарая
Адрес: 68100 Аксарай Турция
E-mail: bilgehanaytac@aksaray.edu.tr

Приянка Трипатхи
Доцент кафедры английского языка Факультета гуманитарных и социальных наук Индийский институт технологий, Патна
Адрес: Комната № 216, 2 этаж, корпус 1 Бихта, Патна — 801103 Бихар, Индия
E-mail: priyankatripathi@iitp.ac.in

Апурв Харе
Доцент кафедры управления маркетинга в Индийском институте менеджмента, Нагпур
Адрес: Кампус VNIT, Южная Амбазари-роуд, Нагпур — 440010 Махараштра, Индия
E-mail: apoorv@iimnagpur.ac.in

Субхасис Рай
Профессор в XLRI-Xavier School of Management
Адрес: Р-н Near Circuit House (восток) Джамшедпур — 831001 Джаркханд, Индия
E-mail: subhasis.ray@xlri.ac.in

В рамках данного круглого стола редакторы специального выпуска организовали критический диалог между ведущими учеными в области социологии, критических исследований коммуникации, культурологии, критических исследований в сфере управления и устойчивого развития, чтобы обсудить проблемы, с которыми сталкивается Глобальный Юг при внедрении искусственного интеллекта. В ходе беседы исследуется неравенство возможностей Глобального Севера и Глобального Юга с точки зрения ИИ-нарративов, а также ведется поиск альтернативных форм и методов управления искусственным интеллектом в ситуации Глобального Юга. Крайне критически и с осторожностью относясь к внедрению искусственного интеллекта на Глобальном Юге, участники дискуссии уделяют приоритетное внимание вопросам прав человека, этики, равенства, инклюзивности и устойчивости исторически обособленных сообществ и социальных групп региона.

Ключевые слова: искусственный интеллект, Глобальный Юг, обособленное сообщество, социокультурное развитие, этика, устойчивость

doi: 10.17323/1728-192x-2024-4-196-220

The Idea of Social Space by Pierre Bourdieu[1]

Yurij L. Katchanov
Doctor of Philosophy, Chief Research Fellow at the Institute for Statistical Studies and Economics of Knowledge,
National Research University Higher School of Economics
Address: 20 Myasnitskaya Ul., Moscow, 101000 Russian Federation
E-mail: ykatchanov@hse.ru

Natalia A. Shmatko
Candidate of Sciences, Chief Research Fellow, Department Head at the Institute for Statistical Studies and Economics of Knowledge,
National Research University Higher School of Economics
Address: 20 Myasnitskaya Ul, Moscow, 101000 Russian Federation
E-mail: nshmatko@hse.ru

"Social space" constitutes a key idea of Bourdieu's sociology. According to this concept, social reality in its objectivist aspect can be fundamentally characterized by a multidimensional distribution of agents across social positions. The metatheoretical principle of Bourdieu's sociology is expressed by the thesis that, by postulating the existence of spatial structure, one can overcome the epistemic difficulties associated with sociological concepts that support a flat ontology within the framework of object-oriented realism. Spatial structure can supply solutions to three major challenges: the first one concerns the tendency for the reification of social order and social phenomena, the second — attributivism, which reduces relationships to characteristics, and the third — the predominance of positivist methodology in sociology. Although Bourdieu claims that spatial structure can be interpreted in terms of a social survey as social space, he goes beyond the boundaries of empirical generalizations. The paper examines the ontology of the idea of "social space". The authors argue that the postulated spatial structure, which calls for realistic treatment by sociologists, constitutes a metaphysical entity that does not lend itself to complete scientific confirmation. To maintain relative proximity of this metaphysical postulate to scientific practice, Bourdieu identifies the fundamental structure of social reality with topological structure, a concept borrowed from mathematics. Topological structure is perceived as a non-phenomenological law that describes relationships between social positions, interpreted as points in social space. The idea of social space claims the priority of topological structure over social positions, which do not exist before, or separately from, the structure. The article considers three layers in the ontology of "social space": mathematical, theoretical, and empirical.

Keywords: social space; Pierre Bourdieu; topological structure; philosophy of science; structural realism; theoretical sociology

Introduction

By placing Pierre Bourdieu's "idea of social space" in the title, we want to emphasize that this sociological approach has not yet been sufficiently established to talk about an established theory. The idea of social space serves as a heuristic principle that helps to systematize and explain disparate sociological concepts. Nevertheless, it is important as

1. The article was prepared in the framework of a research grant funded by the Ministry of Science and Higher Education of the Russian Federation (grant ID: 075-15-2022-325).

it is often the ideas rather than theories that form the basis for the development of sociological trends.

The "social space" is the center of crystallization of an influential research program initiated by Pierre Bourdieu[2]. This makes the study of social space one of the key ideas of modern sociology. The study of the problem field that is formed around the social space is of particular importance, since it can serve as a guideline for social and cultural policy.

Bourdieu's sociology implies a rejection of the ontology of self-existing social agents and a systematic critique of substantialism, understood as naive realism (Bourdieu, Wacquant, 1992). For Bourdieu, social space is a multidimensional distribution of agents according to objective social positions, which makes it possible to scientifically explain and predict practices, strategies, dispositions, representations, etc. (Bourdieu, 1979). Social positions are constructed on the basis of socially effective dispositional properties of agents. A position is characterized by its location in a metric space (the relation of this position to all others) and the measure of socially effective resources or "active properties" that it objectifies. "Social space" is a justification or reification of a logically conceivable topological structure. This structure denotes what is common to all social positions and serves as the environment in which they are implemented. The position becomes a part of the space, its "accident". And the part/whole relationship acts as an abbreviated description of complex processes. Yet there is a conceptual shift: the meaningful hierarchical social order of theories of social stratification and social structure is replaced by a universal form, where spatially similar relationships are fixed. In order to avoid incorrect interpretations, we emphasize that the concept of social space in no way claims to be a "general theory" of society — it can only be considered as a necessary moment of a composite whole. In Bourdieu's sociology, reality is described using the "trialectic" of symbolic, social and physical spaces (Wacquant, 2023: 6–10).

The Bourdieu research program provides neither detailed theoretical explanations of the "social space"[3] nor the axiomatic–logical construction, as is common in formalized theories. Historically, it looks as if Bourdieu analyzed possible empirical manifestations of social space, and then proposed a topological structure that fixed the set-theoretic representation of the characteristics of these manifestations and thus combined them, making the transition from empirical forms of representation to a basic concept (Saint Martin, 2015). The idea of social space arises in the process of a complex interplay of empirical interpretations and *a priori* reflection.

Bourdieu emphasizes that "social space" is defined only within the framework of his other concepts (Bourdieu, Wacquant, 1992: 96), and implicitly — through a relationship with the philosophical (G. Bachelard, G.W. Leibniz, E. Kassirer) and sociological (M.

2. See, e.g., (Atkinson, 2023; Atkinson, Schmitz, 2024; Blasius et al., 2019; Flemmen et al., 2018; Garcia et al., 2023; Jarness, 2018; Reed-Danahay, 2020; Wacquant, 2023).

3. "Bourdieu was insistent that he did not 'do theory'. He constantly warns us against the seductions of pure conceptual disquisition and the dangers of 'theorizing', which so easily veers into scholasticism. … Around 1989, he turned down an invitation by Jeffrey Alexander... to hold a sort of 'world summit' of social theory with Jürgen Habermas, simply because that agenda just did not make sense to him..." (Wacquant, Akçaoğlu, 2017: 40).

Weber, I. Hoffman, E. Durkheim, A. Schutz) ideas of predecessors (Bourdieu, 1989; Vandenberghe, 1999).

The content of the "social space" is not exhausted by its definition as a multidimensional distribution of agents on the basis of socially effective properties. No less significant, if not more, are the results of Bourdieu's vast and in-depth empirical research, which, simultaneously serves as a practical definition, operationalization, and "meaning" of the concept for the working sociologist.

For Bourdieu, the construction of social space was not only the art of multidimensional statistical analysis of heterogeneous empirical data (Blasius, Schmitz, 2014). It was also the result of deep philosophical and sociological reflections that depended on the expected results and each time required creativity and theoretical imagination (Lebaron, 2015). However, from some point on, the "social space" ceased to need independent imagination and understanding. His development largely came down to the systematic application of what was the fruit of research intuition. This has given rise to the illusory representation that there is a universal social space of modern societies (Glevareck, 2023; Hardy, 2013; Vandebroeck, 2018). Nowadays, the "social space" is a quite respectable research program with its own methodology, its own procedures for collecting, processing and interpreting sociological information (Blasius et al., 2019). The transformation of conceptual insight into a recognized research field routinizes scientific breakthrough, constantly pushing back the boundary beyond which discovery begins. This is exactly the effect produced by works (Grenfell, Lebaron, 2014; Robson, Sanders, 2009) and others. If 30–40 years ago, only sharp minds were able to obtain significant scientific results without falling into various kinds of methodological traps, then the efforts of subsequent generations of sociologists have made the "social space" accessible for wider application. Here it should be noted that these efforts have bypassed one essential aspect of the concept of social space, namely, the ontological one (Maton, 2003).

The ontology of sociological theory answers the question: "What is the social reality?" The ontology of the concept of social space is important because all statements about social space can be interpreted as statements about its ontology. The theme of scientific realism is the core of the problem of substantiating sociological knowledge and philosophical reflection on sociology. In our case, the problem is as follows: can the concept of social space be called knowledge, i.e. is it referential to social reality? Or is this concept referential only to sociological practices? The problem of the ontological status of the concept of social space leads us directly to questions about the nature of sociological knowledge and the essence of sociological practices. What should sociology be — just an empirically relevant methodology of interviews, mass surveys and database operations, or something more? Realism insists that science has access to the essential, and not just the phenomenal, level of social reality[4]. Do we have sufficient grounds to believe that the entities [*enses*] postulated by sociology actually exist? How should researchers solve ontological problems right now, when before our eyes the fundamental cognitive agenda is

4. See arguments in favor of unknowability in e.g., (Hanson-Park, 2023; Morganti, 2004).

being replaced by managerial and socio-technological tasks, and there is an active search for new methods of sociological research and explanation?

The purpose of this paper is to analyze Bourdieu's sociological "project", summarizing his ideas and identifying the ontological structure that underlies it. Bourdieu used the concept of "social space" to rely on "hard science" and avoid irrationalism and relativism. The concept of social space emphasized the scientific character of Bourdieu's sociology by giving it a theoretical organization, as opposed to a haphazard set of results, rules and practices. Considering the "social space" as an object of research, we go beyond it and reveal that its own elements are part of a broader concept of structural realism.

Social Space as a Theoretical Concept

The idea of social space is a theoretical concept. Speaking of an idea, we draw a distinction between the empirical and the structural research of the social space. That is, the structural content is significant and systematically provides not only a method, but also an inner theoretical form for Bourdieu's sociology — a form that asserts its own universality.

Bourdieu follows E. Cassirer in perceiving space as a "possibility of togetherness" ("Möglichkeit des Beisammen") (Cassirer, 2010: 487), as a "space of idea" (Cassirer, 2010: 488) that can be studied through the order of relationships rather than as an abstract description borrowed from mathematics. To comprehend social space means to develop a relational conceptual framework through a system of assumptions, hypotheses, and postulates. To achieve that, Bourdieu equates social space to topological space, creating a scientifically founded and respectable insight that asserts: the properties and dynamics of social phenomena are conditioned by the underlying topological relationships (Bourdieu, 1985: 723–726).

At the same time, Bourdieu does not propose a mathematical model, but states that social space is a positional attribute of the social reality and that there is a mathematical form that captures the inherent spatiality of social reality — a form that allows for the empirical diversity of elementary differences (Bourdieu, 1997: 162). By focusing on social space, Bourdieu avoids offering a straightforward model of social reality but introduces a mathematical structure that serves as the basis for its description and explanation. N. Bourbaki gave the first comprehensive definition of "structure": "It can now be made clear what is to be understood, in general, by a… structure. The common character of the different concepts designated by this generic name, is that they can be applied to sets of elements whose nature has not been specified; to define a structure, one takes as given one or several relations, into which these elements enter… then one postulates that the given relation, or relations, satisfy certain conditions (which are explicitly stated and which are the axioms of the structure under consideration). To set up the axiomatic theory of a given structure, amounts to the deduction of the logical consequences of the axioms of the structure, excluding every other hypothesis on the elements under consideration…" (Bourbaki, 1950: 225–226). If the notion of structure for social sciences

is to hold any water, it should be based on this definition that captures the fundamental intuition underlying it.

In that way social space is an ensemble of sociological *enses* that possesses a relational structure. social space does not provide a direct representation of social phenomena, but rather characterizes them through positions and distances (Bourdieu, 1979),. Seemingly obvious, the conceptual simplicity masks the philosophical ideas that underlie the concept of "social space". An apparently intuitive methodology hides a non-trivial ontology: behind the method there is a topological structure that "encodes" social structure.

From the philosophical viewpoint, Bourdieu's approach to social space may be labelled as "rationalistic realism" (Vandenberghe, 1999), "epistemological realism" (Lahire, 2023), or "structural realism"[5], with E. Cassirer among the forerunners of the latter (in the form of conceptual structuralism: see (Ferreirós, 2023)). Bourdieu drew the following conclusions about the nature of theoretical knowledge from the views of Cassirer (French, 2014: 99):

- relations with known logical and mathematical properties conceptually precede *relata* (objects related to each other);
- scientific objectivity shifts from objects to regularities.

Bourbaki's mathematical structuralism was not the sole source of the idea of social space. Bourdieu was also deeply influenced by the French school of structuralism and semiotics, which started with the object as an organic whole, and "emphasized the irreducible relations linking elements together" (Aubin, 1997: 311). The role of structure in Bourdieu's sociology is seen as "unity of diverse aspects" (see (Bourdieu, Wacquant, 1992: 17–19)):

- structure in social science is not a primitive element; it is defined through an ensemble of relations but it is qualitatively different from any one of them and it forms their totality;
- the content of each relation depends on the content of the structure in which it is incorporated;
- structure does not stand in opposition to its elements, since not only are the elements determined by their structure, but the structure is also determined by its elements;
- the relations that make up the structure do not form a chaotic conglomerate, but have a specific order (or orders).

5. Scientific realism is defined by its core assumption that theoretical knowledge is possible and offers an accurate description of reality (Chakravartty, 2007; Rowbottom, 2019). At the same time, no logical contradiction arises between adherence to scientific realism and the deviation of metaphysical realism (Alai, 2023; Corti, 2023), i.e., the thesis about structured reality that exists independently of reason. Simply put, structural realism is a minimalistic version of scientific realism: it solely argues for the existence of a mathematical or structural content of scientific theories. For an excellent and in-depth overview of structural realism see (Ladyman, 2023). While heavily focused on physics, structural realism can in principle be used *mutatis mutandis* to explore the problems of social science (Kincaid, 2008; Porpora, 2022), although opposing views have been voiced (Lyre, 2013; Tulodziecki, 2016).

Bourdieu insisted that sociology, in its objectivist mode, is a social topology and that thinking in the terms of social space is thinking in terms of topology[6] (Bourdieu, 1989: 16). Based on *a priori* considerations and verified by sociological experience, "social space" is introduced as a topological structure defined on some carrier. The "topological" definition ensures intelligibility. "Social space" is a theoretical concept, and a theoretical concept generates sociological knowledge by means of deductive reasoning without directly referring to empirical data. This constitutes its *differentia specifica*. Why is this possible? Classical German philosophy states that the basis of the hypothetical-deductive development of theoretical knowledge is its definitive attributes: apodicticity, unconditional universality and apriority. What necessary and universal knowledge is contained *a priori* in the definition of social space? What kind of *a priori* knowledge can sociology use in principle? The answer is obvious: mathematical knowledge. Similarly, the concept of topological structure makes Bourdieu's "social space" a theoretical concept. Topological structure forms a non-empirical (*a priori*) aspect of the concept of social space. The correct use of the concept of social space (and not metaphorical, as, e.g. in (Liu, 2021)), implies that the *essentia* of social reality is adequately described by mathematical structures. These structures are built in accordance with logical rules (independent of science) in such a way that their features are discovered rather than invented.

Bourdieu assumes without proof that social reality can be mapped into a mathematical construct such as (metric) space. It is not about mathematically formulated sociological theory. It means that sociological research projects an object constructed by sociologists onto a certain metric space, which, in turn, requires a meaningful sociological interpretation.

At the same time, the mathematical content of social space is not just a representation of nominalist (sociological) content — they cannot be separated (cf. (Leng, 2020)). Bourdieu compares social reality with mathematical space based on their common attribute of "structurality". In other words, the connection of social reality with the mathematical concept of space is motivated by the fact that reality is considered from the viewpoint of structure. Social phenomena become part of spatial order insofar as they are structured, and structured by sociological research.

Bourdieu seeks a replacement for the theories of a "social structure" that contains all the specific sociological structures. Therefore, the concept of social space, while retaining important structural information, is used not to construct a theory, but to highlight a theoretical object. Topological structure does not become the main tool of sociological research, but it is used in fundamental reasoning. This means that constructive postulates and statements about social space correlate with the existential propositions of topological structure only at the level of mathematical ontology, while the construction of specific sociological entities (social positions, capital, etc.) takes place at the empirical level. The Ontological commitments of the concept of social space do not conflict with its epistemic commitments.

6. See a detailed analysis of Bourdieu's topological "dimension" in (Wacquant, 2023: 37–40).

Bourdieu's notion of social space is based on a constructive scheme of structure, but it requires resources that are not directly linked to mathematically defined structure. In our case, the structure is one or more sets that are in external relations, which are characterized by axioms[7] defined in the terms of set theory (Bourbaki, 2006: E IV.4–5). However, not all of Bourdieu's statements are conditioned by structural representation or can be explained in structural terms: structure is not enough to identify nonstructural properties, for example, monadic relations (attributes) or internal states (Bourdieu, 1979).

Topological Structure

The construction of a social space, the implementation and development of sociological research within the framework of this program are impossible without mathematics encoding sociological entities. But mathematics, which has become a factor and a means of studying social space, is not always directly related to topology. More often it is about probability theory and statistics. Nevertheless, the topological structure is fundamentally important for the idea of social space. It is not a combination of propositions like set theory, which, being a consistent syntactic construction, allows to make plausible statements about social reality without referring to data. No, the topological structure does not guarantee the logical form of possible empirical propositions. It acts as a prerequisite and a condition — one of the conceptual forms of sociological experience. This is possible because the concept of topological structure is directly included in the construction of the social space. In an idealized description, topological structure is identified with the structure of social reality, whereas social space is considered isomorphic to the mathematical one in its construction.

To clarify our subsequent argument, we refer to the set-theoretic definition of structure[8]. So, structure S comprises:

- the definition area — a nonempty set of objects U,
- a nonempty ordered set of R relations on U.

Structure is commonly treated as an ordered pair: $S = [U, R]$ (Krause, Arenhart, 2016). In these notations, there are two structures $S_1 = [U_1, R_1]$ and $S_2 = [U_2, R_2]$ if there is a one-to-one mapping $\varphi: U_1 \to U_2$ that φ preserves the system of relations of structures S_1, S_2 in the following sense: for all relations $r_1 \in R_1$ and $r_j \in R_2$ indexed set $(a_1, \ldots, a_n)$ from U_1 satisfies the relation r_i, if indexed set $(b_1 = \varphi(a_1), \ldots, b_n) = \varphi(a_n))$ in U_2 satisfies r_j, where $r_i \in R_1$ is a relation corresponding to $r_j \in R_2$ (relations r_i and r_j have the same index in ordered sets of relationships R_1 and R_2). In that case, φ is called isomorphism (da Costa, Rodrigues, 2007). Isomorphism and homology are often used to analyze both

7. Axioms limit the notion of social space. From the perspective of empirical science, they can be considered hypotheses.

8. N. Bourbaki's "structures" can facilitate effective articulation and formalization of scientific theories (Bolinger, 2015: 59–68 u. ff.), although other approaches to the representation of structure are also possible (see, e.g., (Bueno, 2021; Frigg, Nguyen, 2020)).

scientific knowledge (Krause, Arenhart, 2016: 15–17), and social space (see, e.g., (Atkinson, 2021; Atkinson, Schmitz, 2024)).

According to structuralism, "concepts constitute the core of mathematics and concepts are captured by the axiomatic method" (Marquis, 2020: 44). A topological structure on a set X is a family Ω of subsets of X that satisfy the axioms of topological structures (Singh, 2019: 7). Namely, a family Ω of subsets of a set X forms a topological structure if it contains an empty set í, X itself, each sum of any number, and each intersection of the total number of its subsets. Sets of Ω are typically called open sets. The pair (X, Ω) are referred to as topological space. The elements (X, Ω) are called points, and the set X — the carrier of the topological space. Topological space is a set with a defined family of all its open subsets. Whereas metric space can be viewed as an axiomatization of the concept of the proximity of points, topological space seems to axiomatize a broader concept of the proximity of a point to a set. Any metric or ordinal structure on *X* produces a topological structure. Therefore, we shall only deal with topological structure from this point onward.

If we deem the structure of social space to be topological, we accept that, at the level of mathematical ontology, all sociological entities can be modeled on a carrier set in terms of the primitive relation of belonging[9]. This provides a useful theoretical framework. The carrier set of social space must satisfy two conditions:

- the carrier of a social space is a connected set (within the limits of certain relations);
- the social differences between elements of the carrier are subject to certain statistical regularities.

When those conditions are met, the construction of social space is rewarded with a sociologically meaningful result.

Across various social spaces (different countries and/or regions of social phenomena), empirical referents of the points do not coincide. However, topological structure as an ontological invariant may endure regardless of the specific elements in those spaces. This became possible because an abstract structure is not implicitly reliant on objects. Elements that form the domain are devoid of any primary (inherent) properties, and relations form tuples (ordered sets) of elements that do not lend themselves to intensional interpretation. A relation is defined extensionally, through the set of tuples to which it can apply, since the mathematical characteristics of a relation are determined only by its extensional. This narrows dramatically the set of intensions that such a relation can imply. The extensional interpretation of structure, however, does not deprive it of its meaning: elements comprising X are interdependent and can be exhaustively described in terms of structure.

Further, it is impossible to establish a one-to-one correspondence between the extensional and the intensional of a structural relationship. So that it is allowed, there are the relations whose extensionals coincide while their intensionals differ. The extension-

9. The language of set theory consists of first-order logic and the only non-logical symbol ∈.

al quality of abstract structure and its unwillingness to explicitly consider intensions is sometimes deemed as its disadvantage (Arenhart, Bueno, 2015; French, 2023). This, however, underscores the generality of abstract structure, which remains *de facto* independent from realizations.

Social space explains the subject of research, written as points, by placing them in an encompassing topological structure. Topological (metric) relations of points express the combination of differences between positions (Bourdieu, 1998: 32) and reveal their external inherent properties. Social space allows for an empirical method of establishing the equivalence of social differences between (individual and collective) agents and the distances between them. Such a principle is largely consistent with the popular spatial metaphor of sociological experience (Bourdieu, 1979). In other words, the differences between sociological entities, for example, (xDy), (xDz), (yDz), are transformed into the distances between them q(x, y), q(x, z), q(y, z), (for example, see (Le Roux, Rouanet, 2010: 35, 37)). Here we face a contradiction: distance is non-negative, symmetric, transitive and self-similar. By contrast, difference is devoid of any such useful mathematical properties: it lacks the important property of transitivity $(xDy) \wedge (yDz) \not\Rightarrow (xDz)$, being merely anti-reflexive $\neg(xDx)$ and symmetric $(xDy) \Rightarrow (yDz)$. To translate differences to distances correctly, the researcher is forced to introduce certain additional conditions, which may lead to conceptual aberrations. Moreover, not every empirical study allows for determining the distances between objects based on the available data, leaving the researcher to make do with differences[10]. Thus, it is often impossible to plot a metric space on a graph (extracted, e.g., from a social network). Therefore, the more general concept of topological space comes into play.

Speaking of social space, Bourdieu assumes by default that for each pair of different points there is a neighborhood comprising exactly one of the points. This means that, in principle, we can separate any two subjects of sociological research or parts of a single subject, and the whole is the sum of its parts. In topology, this property is referred to as separability (Singh, 2019: 89–90). Separability of social space may signify a persistence of social differences.

As a result of an axiomatic choice that does not require an external justification, topological structure acts as a constructive definition for the theoretical subject — the idea of social space. It also functions as a mathematical condition for such an idea. The connection between Bourdieu's sociology and topological structure reveals itself as a mathematical level of ontology.

Mathematical ontology and sociological ontology, however, are far from identical. It is obvious that social reality as such cannot consist of an ordered set of tuples devoid of primary properties. Sociology here acts as a meta-ontology of a topological structure. Bourdieu's sociological statements appear excessive from the viewpoint of topology. The fact that sociological ideas of social space extend beyond topological structure creates a premise and a condition for research. Social reality is not mathematical, but can be

10. See, e.g., (Katchanov et al., 2019).

described in mathematical terms. In sociology, it is possible to identify specific structures in which agents and relationships are empirically interpreted, i.e. sociologically adequate. Basic structural relations (for example, "to be more than") in social science have intensionals, and this allows us to explain the sociological content of the spatial structure (Belardinelli, 2023). Speaking of structure that constitutes social space, Bourdieu undoubtedly believed that it is isomorphic to the structure of social reality (Bourdieu, 1989: 16–17).

The logical simplicity of Bourdieu's concept presents the concept of social space as a self-evident general term that captures the available volume of empirical facts. As a result, statistical regularities appear as a logical necessity. Postulation of a topological structure underpinning social space may signify an invitation to research and not a mere doctrinal gesture. This structure is not closed. It can be modified and interpreted depending on the research subject and the abilities of the sociologist[11]. Topological structure and spatial relations are fundamental and become *ens realissimum* (Bourdieu, 1998:31) in the interval of idealization assumed by Bourdieu for the plane of objectivist ontology. Outside this plane, points (positions) in social space can have constituents and primary (internal) properties. In general, social positions represent complex entities, while topological structure should be construed as external relations that are not fully dependent ontologically on the internal properties of positions (cf. (Paolini Paoletti, 2021)).

The Ontological Problem of Social Space

The ontological problem of the notion of social space can be formulated as follows:

- the term "social space" may correspond to referents in (an independent from the sociological research) reality or constitutes a description of data of sociological experience data[12];
- the concept of social space is either true (in the sense of the correspondence theory of truth — the core of realism (Leeds, 2007; Psillos, 1999)), or it is just a tool for describing sociological experience.

When study is oriented towards social reality rather than its own products, it is bound to encounter controversy. Therefore, sociologists who do not limit themselves with hypothetical constructs hardly ever engage in protracted chains of deductive reasoning[13] in an attempt to avoid controversy.

11. Topological structure, with its minimal sociological content, makes the theoretical concept of social space invariant to any social-historical context. See, e.g., (Barth et al., 2023; Guy, 2018; Lu, Fan, Fu, 2021).

12. Broadly speaking, the realism of "social space" reveals itself as a thesis that this notion offers a way for the objective study of reality. The thesis that the postulated general concept actually exists is a special case of the previous thesis (see (Musgrave, 2017)).

13. "Social space tends to be translated, with more or less distortion, into physical space, in the form of a certain arrangement of agents and properties. It follows that all the divisions and distinctions of social space... are really and symbolically expressed in physical space appropriated as reified social space..." (Bourdieu, 2000: 134).

It is always possible to propose several alternative ontologies of social space, and there is no fundamental way to make the single correct choice. From the semantical viewpoint (Suppe, 1989), the resulting empirical underdeterminedness denotes the fact that Bourdieu's concept lacks a comprehensive theoretical definition of the ontology of social space:

- theoretical fragments of Bourdieu's writings do not provide a nomenclature of social position;
- it is argued that social positions are objectified by "classes on paper", i.e. statistically constructed research subjects;
- empirical studies produced by adherents of the research program established around social space tend to use different and variously constructed social positions.

The definition of the concept of social space is formulated in such a way that its operationalization in empirical data constitutes a separate complex problem. The concept alone cannot determine which objective social positions exist — this is a task for empirical research. Social space demonstrates ontological commitment to sociological entities only if they exist within all of its possible ontologies. Such entities, however, do not exist. Social space, on the one hand, offers a conceptual framework and explanatory principle. On the other hand, it is impossible to unequivocally derive ontological consequences from such a scheme.

Nonetheless, we deem the empirical underdeterminedness of the ontology of social space in Bourdieu's sociology an asset rather than a flaw. To be able to use this asset efficiently in future scientific practice, we shall have to shift our focus from spatial intuition and comprehension to a formal definition. The confirmation of the social space ontology does not solely depend on empirical consequences. Instead, a choice between empirically equivalent ontologies can be based on theoretical virtues such as explanatory power, unifying power, the ability to generate novel and testable predictions, plausibility, consistency with other accepted theories and background beliefs, simplicity, etc. (Alai, 2019).

Formalization becomes necessary since the loose interpretation of social space as a correspondence established between empirical statements and their generalizations, on the one hand, and abstract structure, on the other, prevents researchers from obtaining systematic results. Thus, sociology is forced to operate with discrete if brilliant insights. Intuition is all good, but it cannot be controlled. While formalization ensures the reproducibility of sociological knowledge. Sociologists need logical purposes in order to develop a conceptual framework for understanding social reality. Bourdieu succeeded in attaching scientific aura to sociological thought by giving it topological structure

The underdeterminedness of the ontological status of social space prompts us to make a choice:

- To consider this concept as a grounded model for empirical research and an outline for a possible future theory that will clarify Bourdieu's ontology.
- To reject the idea that the ontology of social space has an epistemic value. This position will be supported by scientists who prefer sociological discourse *per se*

and replace consistent and binding theories with collections of witty aphorisms or scandalous observations, prone to ambiguous if not conflicting interpretations.

The problem of the ontology of social space is split into two:

- What is the ontological status of the topological structure? Does it exist on its own?
- What is the ontological status of individual sociological entities, positions of the social space?

Bourdieu argues that social space is the first and the last reality because it determines all the ideas that agents may have about it (Bourdieu, 1998: 13). Social space is as real as physical space: upward movement in the former requires effort, work, and time and is inevitably accompanied by the acquisition of marks, or stigmata, of this effort (Bourdieu, 1985: 725–726). According to Bourdieu, social space is the very "social reality" that Durkheim wrote about, i.e. a set of objective relations between positions that cannot be reduced to interactions (Bourdieu, 1989: 16–17). Social space tends to be more or less directly objectified in physical space. Its topological structure generally corresponds to the order of agents' coexistence in inhabited (or appropriated) space (Bourdieu, 2018: 107).

Social space, therefore, is not an empty stage where processes unfold and social agents should only play certain roles. Social space is it a mere absence of something. It constitutes a certain sociological entity *per se*. Agents cannot "teleport" but must brave the journey through social space.

Social space is not a reality hiding behind the phenomena of the social world. Instead, it is an ontologically binding structure and form, where reality reveals itself to sociological research. The conception of social space is the conception that the research subject contains not only individual empirical entities and their properties but also something else. The notion of social space assumes that, alongside agents, social differences and the transactional relations between them, there exists an objective topological structure[14]. This structure cannot be ignored, despite being a product of contingent social-historical processes (Atkinson, 2021).

Here, we are confronted with an obvious discrepancy:

- on the one hand, objective social positions are epistemic entities, i.e. they are provided irrespective of sociological research (Bourdieu, 1984: 34-45);
- on the other hand, social space with sociological constructs as its points determines the perceptions, thinking, and practices of indisputably existent individual agents. This space is just as real as its physical counterpart.

14. The adherence of topological structure to ontology is determined by methodological schematization and interpretational perspective, which are recognized and methodologically implied by the notion of social space (Lenk, 2017). "Working sociologists necessarily operate with a realist ontology It therefore follows that the success of the critique of reism [Reism is a philosophical doctrine that argues the category of material things to be the only ontological category — Authors' note] is effectively guaranteed in advance, but what is still more troubling is that the critique of reism has the potential to put into question the very possibility of sociology" (Vandenberghe, 2008: 11).

To ease the tension, we shall identify the nature of social space, whose reality we maintain. Since objective social positions are referential only with respect to sociological experience, the best solution is to assert topological structure as something that sociologists regard realistically. Within the framework of the semantic approach, "social space" is a family of models representing the structure of reality (cf. (Suppe, 1989)). We insist that of these models, at least the topological structure is empirically correct and isomorphic to the objective structure of social reality. Social positions in the topological structure constitute a carrier of social space. Thus, we consider statements about certain social positions as statements about a topological structure of a certain type. Adherence to the topological structure makes it possible to preserve referential semantics and assume the position of local realism (Psillos, 1999: Ch. 12) whenever sociology operates with "classes on paper" (statistical constructs devoid of internal order) or epistemic individuals.

The topological structure as a mathematical concept exists *ante rem*, i.e. before and independently of any subjects identified by sociological research. However, social space exemplifies the topological structure solely because the latter models social reality. By contrast, as an empirical exemplification, topological structure exists *in re*.

Does topological structure depend ontologically on its carrier? Topological structure can be defined as an abstract form of social space, such that a single structure is exemplified by several spaces, with social positions varying from one space to another. By refraining at the theoretical level from realistic treatment of sociological entities with an ambiguous ontological status, we cannot make assertions about topological structure as something abstracted from entities. Such a view of the ontology of social space does not help it to converge with the epistemology of social science. Yet neither does it shroud the topological structure in mystery, since topological structure is a system of open neighborhoods, whose nature is irrelevant.

The concept of topological structure is devoid of ontology in the sense that social reality has no concrete topological entities that would be unmetaphorically construed. In this regard, the reality of topological structure is not a definition that could supplement the concept of social space — to use Kantian terminology, it does not constitute a real predicate ("Critique of Pure Reason". B 626). More precisely, the predicate of existence in this case can express the ontological independence of topological structure from the results of the sociological measurement.

The referent of the concept of "social space" is a predetermined objective framework within which social agents think, practice, perceive, etc. The reality of the topological structure is that it structures this framework. By the same token, non-transitive social relationships are immediately mirrored in the relations of proximity between social positions or between agents and positions. The structure of proximity relations is a part of the natural order, being a historical invariant. As a result, topological structure uncovers social reality in its objectivist aspect. Topology describes reality as a space, formulating the purpose and the means for implementation of social trajectory. In the context of social space the term "topological structure" not only characterizes the connections of

many social positions, but also conveys the agents' attitude to social reality. Topology is how reality is revealed to agents. It is an objective horizon where symbolic space is formed.

The concept of topological structure appears to be thematically neutral, and the fact that it is realized in the relations of a family of social positions (statistically constructed sociological entities) does nothing to invalidate its mathematical status. However, mathematics is not ontologically neutral: subjected to relevant empirical interpretation, it provides the means for accessing social space. The fact that statistical or epistemic social agents do not possess the same ontological status as topological structure is of no concern. According to Bourdieu, the study of agents obscures or replaces the study of complex multilevel sets of relations. Social positions can be empirically interpreted, but they remain ontological derivatives of topological structure. In the idealization interval stipulated by the concept of social space, topological structure, holds precedence, while all sociological entities should be interpreted in structural terms (a viewpoint argued by structural realism (French, Ladyman, 2011)).

Why does topological structure ontologically precede sociological entities? It is possible that each entity in the social space stands in a structural relationship with any other entity (Morganti, 2019). Therefore, relations and entities mutually support and condition each other. The common interconnectedness of entities in social space and their complete immersion in the network of structural relations, as argued by Bourdieu (Bourdieu, Wacquant, 1992: 16), supports the thesis of the ontological primacy of topological structure although the holistic thesis of the global interconnectedness can generate logical-philosophical problems of its own (Busse, 2023; Swiderski, 2022).

Layers of Ontology

The metatheory of social space calls for formalization that would differ from what is already in use by sociologists in a direct or implicit way. We shall take a close look at the following essential characteristics:

- mathematical ontology,
- the ontology of the theoretical concept of social space,
- empirical statements.

These three layers of ontology are typical of contemporary scientific theory (Krause, Arenhart, 2016). Bourdieusian sociology views topological structure as a mathematical "*materia prima*", from which social space is derived. In turn, the topological structure is not a primitive. It is constructed on the base of set theory, which proceeds from deeper principles forming the foundations of mathematics. In fact, we recognize the existence of extra-empirical principles that can only be expressed mathematically. Thus, we overcome the challenges of the irreducibility of social space to "datasets" and the presence of ineliminable super-empirical content.

Ontological commitment to social space implies realism regarding the type of mathematics required for the identification of topological structure. What can we know about

social space — or rather, topological structure — from the viewpoint of the set-theoretic interpretation of structure? Social space appears to us as an ordered family of sets. Set theory allows for social space (topological structure) to be viewed as comprised of separate scientific entities and determined by its internal structure of belonging (Barton, Friedman, 2019). Based on set theory, topology offers the possibility to study social positions in their various combinations. At the same time, a topological structure considered together with its carrier can be identified as a scientific entity with its own attributes. This renders topological structure or social space significant as an integral subject of sociological research. Reification, or hypostatization, constitutes a fundamental logical and semiotic phenomenon. To study topological structure means to identify it within the framework of first-order logic by asserting predicates and quantifying those assertions. Topological structure is an abstraction devoid of physical existence but possessing internal connections with well-defined subjects of sociological research. The manner of existence of topological structure differs from that of physical objects. It remains objective in the sense that it is intersubjective to the highest degree.

The subject of the study requires clearly defined identification conditions in the ensemble of vertical relations "theory — data" in order to consider it a family of social positions. Reliance on the theoretical concept of social space cannot yield such a clear identification criterion. By limiting ontological commitment to topological structure, we ensure that our arguments are supported by the theoretical concept of social space. Thus, we can maintain scientific realism in the face of methodologically controversial "standard" social positions, which can be construed as ontologically irrelevant scientific entities (Bourdieu, 1989: 19–20).

Our approach can be summarized as follows: the social space is empirically underdetermined, since there is no uniform and complete list of positions and socially effective dispositional properties for its construction. Moreover, its practical implementation allows for countless models with different properties.

The idea of social space constitutes an ontology. This idea can be perceived uncritically: although initially assumed as a hypothesis, social space has properties that are nonrandom and necessary. At the same time, the referentiality of the theoretical concept of social space is an implication of its truth or plausibility, since the selection of the referent is conditioned by the description.

However, as soon as we use statistical methods of dimension reduction and classification, we arrive at a less definite structure — a two-dimensional projection and clusters — we obtain hypothetical results that can be challenged. The properties of social space as a whole constitute topological invariants. These properties vanish, however, once we achieve the transition from the totality of distances between all positions to a visualization of topological structure by means of statistical algorithms. A two- or three-dimensional visualization, common in scholarly publications, is not social space as such, but a projective representation thereof.

At the mathematical level of the ontology of social space, topological structure can be treated as an empirically unrepresented concept. At the sociological level, however, it

becomes represented by a system of spatial relations of a set of social positions. While at the first level we are dealing with a topological structure postulated by means of formal axioms, at the second level sociology operates with informal concepts of social positions, which are constructively described using empirical data. Yet the mathematical level of the ontology of social space enables us to convert informal empirical results into formal statements. E.g., the two-dimensional mapping of social space, so popular among sociologists (Blasius et al., 2019), is a mere approximation of topological structure. Nonetheless, a visualization of social space using the Euclidean plane introduces mathematical structure with its points and real numbers (distances, angles, etc.).

We can speak of a certain pluralism: "social space" depicts reality using two diverse but overlapping ontological schemes — theoretical and empirical (cf. (Glick, 2021)). Although these schemes map social reality in slightly different ways, they are coherent. In general, the ontology of social space can be captured as an interference of theoretical and empirical ontologies.

In the process of sociological research, the theoretical concept of social space collapses — it turns into a set of factual statements. Although charged theoretically, they no longer represent *a priori* universal apodictic knowledge. Instead, they become empirical statements. At the empirical level, topological structure provides an epistemic resource and does not signify a sociologist's ontological commitment. Topological structure evolves into a plane image constructed with the use of mathematical statistics and accompanied by its verbal description (Lebaron, 2021).

What is "social space," then? Is it a substantiation of topological structure reflecting the essential characteristic of reality, or a geometric metaphor — a mathematical fiction — whose explanatory power should be questioned? Although mathematical tools sometimes help sociologists, a mathematical concept can never replace sociological knowledge. Mathematical formalism can facilitate the progress in sociological description and explanation only to the extent that it is strongly related to unformalized ideas of social reality. For Bourdieu, topological structure is not a mathematical construct that operates separately from sociological statements. On the contrary, he views it as a cornerstone of a sociological concept that ensures organic incorporation of diverse empirical elements into a complete whole.

Therefore, the idea of social space is introduced by means of the axioms of topological structure, whereas the idea of topological structure, in turn, is bestowed with sociological content by the concept of social space. Thus, we encounter *circulus in probando*: the concept of social space needs the concept of a topological structure, which cannot be sociologically defined without the concept of social space. Within the framework of sociology, "social space" is defined by a "topological structure", and "topological structure" is defined by a "social space". If the notion of social space could be defined independently of the notion of topological structure, this type of circular reasoning would not occur.

Let us clarify the point made above. The concept of a topological structure assumes that every element , belonging to an open set (which in effect means that satisfies a set of conditions) has a property . However, the sociologist defines the topological structure of

on the set empirically — by taking an arbitrary family of sets and declaring it a prebase of topology. To be more specific, this family is deemed random in mathematics, but sociology views it as justified. There is an obvious isolation: the topological structure of is uniquely determined by the researcher's sociological choice. It does not have an independent sociological content, since the topological structure in sociological research plays the role of a form organizing the results of empirical research external to it. Topological structure may be built on any data prepared for the construction of social space (so long as conformity to its axioms is ensured). The result of this construction will be interpreted as social space. In other words, neither of the two terms can boast an independent sociological definition. However, here we are dealing not with paralogism, but with the pragmatic specificity of the axiomatic method: "social space" and "topological structure" are introduced as theoretical concepts undefined in another — more general — sociological conception. This problem is removed by an explicit reference to the ineliminable gap between the mathematical and theoretical layers of the ontology of a sociological concept. At the same time, while studying a social space, the boundaries of mathematics and sociology intersect. In each specific case, it is possible to point out a more or less pronounced effect of preliminarily nominalized mathematics on the findings of sociological research: they share structural similarity.

A sociologist can learn a lot, for example, about small entrepreneurs without studying individuals in Forbes' billionaire rankings. However, the relative epistemic autonomy in no way rules out their ontological interconnectedness: small entrepreneurs are only called small because there are also middle-sized and big entrepreneurs. Sociologically meaningful research implies a study of the patterns (common forms abstracted from properties that do not affect the relations between *relata*) of social reality. Social reality asserted in its objective aspect constitutes a system of positions in spatial relations and in the relations of potential changes. Social positions are not singular terms like proper names: they are linked to the structure. Yet the differences between positions in topological structure and objects are relative. It is worth remembering that, from a Bourdieusian perspective, social position is viewed as a group of interconnected properties rather than an object. Social positions are characterized through their relations to each other, while each position is determined by topological structure. It follows that the preferred research strategy is to study small entrepreneurs as a position in social space.

The relativity of social position and existence in the concept of social space does not disprove the fact that each position has a meaning assigned to it by background (social) ontology. Namely, it correlates with families of properties and individual social agents selected on the basis of these families. Background ontology is not an implication of the theory of a certain structure. The difference between social position as a complex scientific entity and social position as a point in space cannot be eliminated. Social position is a relative concept that represents the carriers of spatial relations and co-represents the relations in question.

What are the special characteristics of spatial relations in Bourdieu sociology? He essentially posits intralayer ontological connectivity, which implies interdependence between social positions and their internal properties, on the one hand, and relations

between social positions, on the other (cf. (Ladyman, 2016: 182–183)). It follows that relations in social space (while irreducible to the inner properties of sociological entities) depend on the existence of *relata* (Mohr, 2013). In the empirical layer of ontology, social positions constitute scientific entities that are quantified, endowed with certain properties while being devoid of others, and can form structural relations.

Conclusion

The idea of social space cannot be immediately inferred from data, but it cannot be deduced from *a priori* reflection either. Social space adheres to the methodology of empirical research and is responsible for ontology in sociology. The topological structure is isomorphic to the structure of social reality and is a prerequisite and condition for sociological practices and statements. The concept of social space as it is presented in this paper meets two main requirements for scientific realism: ontological responsibility and scientific methodology (Chakravartty, 2007: Ch. 1).

The significance of topological structure in sociology is determined by its usage. As demonstrated in this paper, the usage can be twofold:

- at the theoretical level, topological structure constitutes the concept of social space as a model of reality;
- at the empirical level, topological structure is a mathematical image of the state of a particular fragment of social reality such that this state is determined not by the values of the quantities, but by the distribution of agents across positions (points or open sets) of social space.

By holding together the theoretical and empirical levels, which support each other, we cover the entire spectrum of meanings of the topological structure.

The perspective of social space proposed by Bourdieu can basically be summarized as follows:

- The topological structure involved as a non-phenomenological law of the ontology of social space corresponds to the structure of social reality.
- The topological structure makes it possible to study not the social phenomena themselves, but their models created by sociologists. Properties described by the axioms of topological structure may appear "thin" from the standpoint of positivist methodology, and as such incompatible with empirical intuition.
- The mathematical concept of a topological structure is projected onto empirical material. Therefore, it can be used to describe and explain facts registered by sociology.
- At the sociological level, the topological structure unites all empirical manifestations of social space, supporting their certain common features and disregarding all others.
- The fundamental use of mathematical space goes beyond the construction of quantitative models and indicates an ontological choice regarding the theoretical image of the studied domain.

- The concept of space used as ontology should be understood not as a mathematical concept *in stricto sensu*, but rather as a record of a set of scientific presentations that are interpreted within an empirical context.
- The concept of social space expresses the structural-relational facts about social reality uncovered by sociology.
- The object of sociological study is considered as a set of social relations together with their carriers. This approach overcomes substantialism.
- Social position is an ordered set of sets of agents. It is constructed on the basis of a combination of dispositional (distributed in social space) properties.
- A dispositional property is treated not as an intrinsic sign of a single agent but as a social relation shared by many agents. Thus, the hypostatization of properties is removed.

Neither philosophy nor mathematics can be used as algorithms that guarantee the production of new sociological knowledge. Sociological statements cannot be logically deduced from philosophical concepts or mathematical theories — their veracity can only be assumed.

The problem of realism may become unsurmountable if discussed at a round table of philosophers. But it loses its metaphysical charge and mathematical aura in the publications of sociologists, gaining its distinct realistic meaning[15]. However, sociologists are not concerned with solving philosophical problems. Instead, they strive to "comprehend" the framework of social reality. Their goal is to create and develop a conceptual framework for thinking about reality that is consistent and agrees with data.

References

Alai M. (2019) The underdetermination of theories and scientific realism. *Axiomathes,* vol. 29, pp. 621–637.

Alai M. (2023) Scientific realism, metaphysical antirealism and the no miracle arguments. *Foundations of Science*, vol. 28, no 1, pp. 377–400.

Arenhart J. R. B., Bueno O. (2015) Structural realism and the nature of structure. *European Journal for Philosophy of Science*, vol. 5, no 1, pp. 111–129.

Atkinson W. (2021) The social space and misrecognition in 21st century France. *The Sociological Review*, vol. 69, no 5, pp. 990–1012.

Atkinson W. (2023) Charting fields and spaces quantitatively: from multiple correspondence analysis to categorical principal components analysis. *Quality & Quantity*. doi: 10.1007/s11135-023-01669-w

15. "...The question about the *general* legitimacy of realist interpretations of well-confirmed mature theories is more radical than the zero level question about the legitimacy of the realist interpretation of a given *individual* theory. The former question is a philosophical question, the latter a scientific one. Clearly, on level zero, i.e., in the scientific context, the general legitimacy of realist interpretations of theories (under appropriate conditions) is taken for granted. ...The general *philosophical* doubt about realist interpretation articulated on level one does not come into play in the scientific practice on level zero" (Hoyningen-Huene, 2018: 7).

Atkinson W., Schmitz A. (2024) The German social space and its homologies: National variation on a basic structure. *Current Sociology*, vol. 72, no 1, pp. 168–191.

Aubin D. (1997) The withering immortality of Nicolas Bourbaki: A cultural connector at the confluence of mathematics, structuralism, and the oulipo in France. *Science in Context,* vol. 10, pp. 297–342.

Barth A., Leßke F., Atakan R., Schmidt M., Scheit *Y*. (2023) Reconstructing a scientist's "space of social relations" via multiple correspondence analysis. *Multivariate Scaling Methods and the Reconstruction of Social Spaces: Papers in Honor of Jörg Blasius Blasius* (eds. A. Barth, F. Leßke, R. Atakan, M. Schmidt, Y. Scheit), Opladen, Berlin, Toronto: Verlag Barbara Budrich, pp. 69–82.

Barton N., Friedman S.-D. (2019) Set theory and structures. *Reflections on the Foundations of Mathematics: Univalent Foundations, Set Theory and General Thoughts* (eds. S. Centrone, D. Kant, D. Sarikaya), Cham: Springer International Publishing, pp. 223–253.

Belardinelli S. (2023) Realism versus relationism. *Methodology of Relational Sociology: Approaches and Analyses* (ed. E. Hałas), Cham: Springer Nature Switzerland, pp. 87–96. 4

Blasius J., Lebaron F., Le Roux B., Schmitz A. (eds.). (2019) *Empirical investigations of social space,* Cham: Springer Cham.

Blasius J., Schmitz A. (2014) Empirical construction of Bourdieu's social space. *Visualization and Verbalization of Data* (eds. J. Blasius, M. Greenacre), Boca Raton, FL, London, New York, NY: Chapman and Hall/CRC, pp. 247–264.

Bolinger R. (2015) *Rekonstruktion und Reduktion physikalischer Theorien: Der Ansatz von Erhard Scheibe an Beispielen aus der Astroteilchenphysik*, Berlin, Boston, MA: De Gruyter.

Bourbaki N. (1950) The architecture of mathematics. *The American Mathematical Monthly,* vol. 57, pp. 221–232.

Bourbaki N. (2006) *Théorie des ensembles*, Berlin, Heidelberg: Springer Berlin Heidelberg.

Bourdieu P. (1979) *La Distinction: Critique sociale du jugement,* Paris: Les Éditions de Minuit.

Bourdieu P. (1984) *Homo academicus,* Paris: Les Éditions de Minuit.

Bourdieu P. (1985) The social space and the genesis of groups. *Social Science Information*, vol. 24, no 2, pp. 195–220.

Bourdieu P. (1989) Social space and symbolic power. *Sociological Theory*, vol. 7, no 1, pp. 14–25.

Bourdieu P. (2000) *Pascalian Meditations,* Stanford, CA: Stanford University Press.

Bourdieu P. (1998) *Practical Reason: On the Theory of Action*, Stanford, CA: Stanford University Press.

Bourdieu P. (2018) Social space and the genesis of appropriated physical space. *International Journal of Urban and Regional Research*, vol. 42, no 1, pp. 106–114.

Bourdieu P., Wacquant L. (1992) *An Invitation to Reflexive Sociology,* Chicago, IL, Cambridge: University of Chicago Press, Polity Press.

Bueno O. (2021) Structural representation and the ontology of models. *Models and Idealizations in Science: Artifactual and Fictional Approaches* (eds. A. Cassini, J. Redmond), Cham: Springer International Publishing, pp. 199–216.

Busse R. (2023) No foundations for metaphysical coherentism. *Philosophical Studies*. doi: 10.1007/s11098-023-02055-w

Cassirer E. (2010) Phänomenologie der Erkenntnis. *Philosophie der symbolischen Formen, Dritter Teil* (ed. J. Clemens), Hamburg: Felix Meiner Verlag.

Chakravartty A. (2007) *A Metaphysics for Scientific Realism: Knowing the Unobservable,* Cambridge: Cambridge University Press.

Corti A. (2023) Scientific realism without reality? What happens when metaphysics is left out. *Foundations of Science*, vol. 28, no 1, pp. 455–475.

da Costa N. C. A., Rodrigues A. A. M. (2007) Definability and invariance. *Studia Logica*, vol. 86, no 1, pp. 1–30.

Ferreirós J. (2023) Conceptual structuralism. *Journal for General Philosophy of Science*, vol. 54, no 1, pp. 125–148.

Flemmen M., Jarness V., Rosenlund L. (2018) Social space and cultural class divisions: the forms of capital and contemporary lifestyle differentiation. *The British Journal of Sociology*, vol. 69, no 1, pp. 124–153.

French S. (2014) *The Structure of the World: Metaphysics and Representation,* Oxford: Oxford University Press.

French S. (2023) Quasi-structural realism. *Non-Reflexive Logics, Non-Individuals, and the Philosophy of Quantum Mechanics: Essays in Honour of the Philosophy of Décio Krause* (eds. J. R. B. Arenhart, R. W. Arroyo), Cham: Springer International Publishing, pp. 29–43.

French S., Ladyman J. (2011) In defense of ontic structural realism. *Scientific Structuralism* (eds. A. Bokulich, P. Bokulich), Dordrecht: Springer Netherlands, pp. 25–42.

Frigg R., Nguyen J. (2020) *Modelling Nature: An Opinionated Introduction to Scientific Representation*. Cham: Springer International Publishing.

Garcia A., Garcia Parpet M.-F., Poupeau F., Pérez A., Rocha M. E. (eds.). (2023) *Bourdieu et les Amériques: Une internationale scientifique: genèse, pratiques et programmes de recherche,* Éditions de l'IHEAL.

Glevarec H. (2023) Social space as a theory of society: Scientific arguments regarding the figuration of the social in Bourdieu's Distinction. *Sociology*, vol. 57, no 1, pp. 96–119.

Glick D. (2021) Pluralist structural realism: The best of both worlds? *Synthese*, vol. 198, no 5, pp. 4145–4166.

Grenfell M., Lebaron F. (eds.). (2014) *Bourdieu and Data Analysis: Methodological Principles and Practice,* Oxford: Peter Lang Verlag.

Guy J.-S. (2018) Bourdieu in hyperspace: from social topology to the space of flows. *International Review of Sociology*, vol. 28, no 3, pp. 510–523.

Hanson-Park J. (2023) Structural realism and agnosticism about objects. *Global Philosophy*, vol. 33, no 2, p. 29.

Hardy C. (2013) Social space. *Pierre Bourdieu: Key concepts* (ed. M. Grenfell), London: Routledge, pp. 229–249.

Hoyningen-Huene P. (2018) Are there good arguments against scientific realism? *Philosophy of Science: Between the Natural Sciences, the Social Sciences, and the Humanities* (eds. A. Christian, D. Hommen, N. Retzlaff, G. Schurz), Cham: Springer International Publishing, pp. 3–22.

Jarness V. (2018) Viewpoints and points of view: situating symbolic boundary drawing in social space. *European Societies*, vol. 20, no 3, pp. 503–524.

Katchanov Yu. L., Markova Yu. V., Shmatko N. A. (2019) The distinction machine: Physics journals from the perspective of the Kolmogorov–Smirnov statistic. *Journal of Informetrics*, vol. 13, no 4, #100982.

Kincaid H. (2008) Structural realism and the social sciences. *Philosophy of Science*, vol. 75, no 5, pp. 720–731.

Krause D., Arenhart J. R. (2016) *The Logical Foundations of Scientific Theories: Languages, Structures, and Models,* London: Routledge.

Ladyman J. (2016) The foundations of structuralism and the metaphysics of relations. *The Metaphysics of Relations* (eds. A. Marmodoro, D. Yates), Oxford: Oxford University Press, pp. 177–197.

Ladyman J. (2023) Structural Realism. *The Stanford Encyclopedia of Philosophy* (eds. E. N. Zalta, U. Nodelman), Stanford, CA: Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/sum2023/entries/structural-realism/ (accessed 21 December 2023).

Lahire B. (2023) *Les structures fondamentales des sociétés humaines,* Paris: La Découverte. URL: https://www.calameo.com/read/0002150226ed469b639ff

Lebaron F. (2015) L'espace social. Statistique et analyse géométrique des données dans l'œuvre de Pierre Bourdieu. *La méthodologie de Pierre Bourdieu en action. Espace culturel, espace social et analyse des données* (eds. F. Lebaron, B. Le Roux), Paris: Dunod, pp. 43–58.

Lebaron F. (2021) Geometric data analysis as a tool for reflexivity. *Historical Social Research / Historische Sozialforschung*, vol. 46, no 2, pp. 126–154.

Leeds S. (2007) Correspondence truth and scientific realism. *Philosophical Studies*, vol. 159, no 1, pp. 1–21.

Leng M. (2020) Mathematical explanation doesn't require mathematical truth. *Current Controversies in Philosophy of Science* (eds. S. Dasgupta, R. Dotan, B. Weslake), New York, NY: Routledge, pp. 51–59.

Lenk H. (2017) A scheme-interpretationist and actionistic scientific realism. *Varieties of Scientific Realism: Objectivity and Truth in Science* (ed. E. Agazzi), Cham: Springer International Publishing, pp. 257–276.

Le Roux B., Rouanet H. (2010) Multiple Correspondence Analysis, Los Angeles, CA, London: SAGE Publications, Inc.

Liu S. (2021) Between social spaces. *European Journal of Social Theory*, vol. 24, no 1, pp. 123–139.

Lu P., Fan X., Fu F. (2021) Profile of the super rich in China: A social space analysis. *The British Journal of Sociology*, vol. 72, no 3, pp. 543–565.

Lyre H. (2013) Must structural realism cover the special sciences? *EPSA11 Perspectives and Foundational Problems in Philosophy of Science* (eds. V. Karakostas, D. Dieks), Cham: Springer International Publishing, pp. 383–390.

Marquis J. P. (2020) Forms of structuralism: Bourbaki and the philosophers. *Structures Mères: Semantics, Mathematics, and Cognitive Science* (eds. A. Peruzzi, S. Zipoli Caiani), Cham: Springer International Publishing, pp. 37–57.

Maton K. (2003) Reflexivity, relationism, & research: Pierre Bourdieu and the epistemic conditions of social scientific knowledge. *Space and Culture*, vol. 6, no 1, pp. 52–65.

Mohr J. W. (2013) Bourdieu's relational method in theory and in practice: From fields and capitals to networks and institutions (and back again). *Applying Relational Sociology: Relations, Networks, and Society* (eds. F. Dépelteau, C. Powell), New York, NY: Palgrave Macmillan US, pp. 101–135.

Morganti M. (2004) On the preferability of epistemic structural realism. *Synthese*, vol. 142, no 1, pp. 81–107.

Morganti M. (2019) From ontic structural realism to metaphysical coherentism. *European Journal for Philosophy of Science*, vol. 9, no 1, # 7.

Musgrave A. (2017) Strict empiricism versus explanation in science. *Varieties of Scientific Realism: Objectivity and Truth in Science* (ed. E. Agazzi), Cham: Springer International Publishing, pp. 71–93.

Paolini Paoletti M. (2021) Structures as relations. *Synthese*, vol. 198, no 11, pp. 2671–2690.

Porpora D. V. (2022) The metaphysical issues in the social sciences and how social scientists debate them. *Synthese*, vol. 200, no 6, # 501.

Psillos S. (1999) *Scientific Realism: How Science Tracks Truth,* London: Routledge.

Reed-Danahay D. (2020) *Bourdieu and Social Space: Mobilities, Trajectories, Emplacements,* New York, NY, Oxford: Berghahn.

Robson K., Sanders C. (eds.) (2009) *Quantifying Theory: Pierre Bourdieu*, Dordrecht: Springer Netherlands.

Rowbottom D. P. (2019) Scientific realism: what it is, the contemporary debate, and new directions. *Synthese*, vol. 196, no 2.

Saint Martin M. de (2015) From "Anatomie du goût" to "La Distinction". Attempting to construct the social space. Some markers for the history of the research. *The Routledge Companion to Bourdieu's Distinction* (eds. P. Coulangeon, J. Duval), London: Routledge, pp. 15–28.

Singh T. B. (2019) *Introduction to Topology*. Singapore: Springer Singapore.

Suppe F. (1989) *The Semantic Conception of Theories and Scientific Realism,* Urbana and Chicago, IL: University of Illinois Press.

Swiderski J. (2022) Varieties of metaphysical coherentism. *Erkenntnis*.

Tulodziecki D. (2016) Structural realism beyond physics. *Studies in History and Philosophy of Science Part A*, vol. 59, pp. 106–114.

Vandebroeck D. (2018) Toward a European social topography: the contemporary relevance of Pierre Bourdieu's concept of "social space". *European Societies*, vol. 20, no 3, pp. 359–374.
Vandenberghe F. (1999) "The real is relational": An epistemological analysis of Pierre Bourdieu's generative structuralism. *Sociological Theory*, vol. 17, no 1, pp. 32–67.
Vandenberghe F. (2008) *A Philosophical History of German Sociology,* London, New York, NY: Routledge.
Wacquant L. (2023) *Bourdieu in the City: Challenging Urban Theory,* Cambridge: Polity Press.
Wacquant L., Akçaoğlu A. (2017) Practice and symbolic power in Bourdieu: The view from Berkeley. *Journal of Classical Sociology*, vol. 17, no 1, pp. 37–51.

Идея социального пространства Пьера Бурдьё[16]

Юрий Качанов
Доктор философских наук, главный научный сотрудник Института статистических исследований и экономики знаний, Национальный исследовательский университет «Высшая школа экономики».
Адрес: ул. Мясницкая, д. 20, Москва,101000 Российская Федерация
E-mail: ykatchanov@hse.ru

Наталья Шматко
Кандидат философских наук, главный научный сотрудник, заведующий отделом Института статистических исследований и экономики знаний, Национальный исследовательский университет «Высшая школа экономики».
Адрес: ул. Мясницкая, д. 20, Москва,101000 Российская Федерация
E-mail: nshmatko@hse.ru

«Социальное пространство» — ключевая идея социологии Пьера Бурдье. Согласно этой идее, социальная действительность в её объективистском аспекте фундаментально характеризуется многомерным распределением агентов по социальным позициям. Метатеоретическим принципом социологии Бурдье является тезис о том, что, постулируя существование пространственной структуры, можно разрешить эпистемические трудности, с которыми сталкиваются социологические концепции, поддерживающие плоскую онтологию в рамках объектно-ориентированного реализма. Пространственная структура используется для ответа на три важных вызова: первый связан с тенденцией субстантивации социального порядка и социальных явлений, второй — с атрибутивизмом, редуцирующим отношения к свойствам, а третий — с доминированием позитивистской методологии в социологии. Хотя Бурдьё декларирует, что пространственная структура может быть интерпретирована в терминах социологического исследования как социальное пространство, он выходит за рамки эмпирических обобщений. В статье изучается онтология идеи социального пространства. Авторы утверждают, что постулируемая пространственная структура, в отношении которой социологи должны быть реалистами, — это метафизическая сущность, которая не может быть полностью подтверждена наукой. Чтобы не уводить метафизический постулат слишком далеко от научной практики, фундаментальная структура социальной действительности отождествляется Бурдьё с топологической

16. Статья подготовлена в рамках гранта, предоставленного Министерством науки и высшего образования Российской Федерации (№ соглашения о предоставлении гранта: 075-15-2022-325).

структурой, заимствованной из математики. Топологическая структура проявляет себя как нефеноменологический закон, описывающий отношения социальных позиций, понимаемых как точки социального пространства. Идея социального пространства утверждает примат топологической структуры над социальными позициями, которые не существуют до или отдельно от существования структуры. В статье рассматриваются три слоя онтологии «социального пространства»: математический, теоретический и эмпирический.

Ключевые слова: социальное пространство, Пьер Бурдье, топологическая структура, философия науки, структурный реализм, теоретическая социология

doi: 10.17323/1728-192x-2024-4-221-248

Pavel Novgorodtsev's Philosophy of Law: "New Liberalism" vs Christian Humanism[1]

Anastasiya Tumanova
Doctor of History, Doctor of Law, Professor,
School of Historical Sciences, Faculty of Humanities, National Research University Higher School of Economics
Address: 20 Myasnitskaya Ul., Moscow, 101000 Russian Federation
E-mail : atumanova@hse.ru

Alexander Safonov
Doctor of Law, Professor, Law Faculty,
National Research University Higher School of Economics (Moscow)
Address: 20 Myasnitskaya Ul., Moscow, 101000 Russian Federation
E-mail : asafonov@hse.ru

The paper characterizes the contribution made by Pavel Novgorodtsev, the Russian philosopher of law, to the establishment of the moral and ethical tradition of law, which, in turn, is the philosophical foundation of the doctrine of new liberalism and human rights. This tradition was influential both in late-imperial Russia and in Europe, to which Novgorodtsev immigrated in 1920 and where he founded the Institute (Faculty) of Law at Charles University (Prague, Czech Republic).

The authors consider Novgorodtsev's legal philosophy as an integral ethical doctrine, taking into account its development in the last period of the scholar's life. For the first time in academic literature, it is argued that Novgorodtsev's views should be interpreted in the context of both the liberal and religious philosophical approaches to the understanding of law, and that these approaches complement each other and have played a significant role in the history of political and legal thought both in Russia and in 20th-century Europe. Novgorodtsev's contribution to intellectual thought in Russia and Europe was in his thoroughgoing criticism of social utopias, particularly, Bolshevism and Communism, which were characterized by their denial of law and legal tradition.

The research is based on a wide range of sources, including P. Novgorodtsev's papers and personal documents, in particular, his correspondence.

Keywords: Russian intellectual history, liberal philosophy, religious philosophy

Problem Setting

In the early 20th century, the Russian philosophy of law produced a constellation of talented scholars. Pavel Ivanovich Novgorodtsev was among the most prominent of them. The head of the Moscow school of legal philosophy and the founder of the Russian liberal tradition in jurisprudence, which was philosophically rich and unwavering in its antipositivist commitment, creator of the Russian neo-idealistic school of the "revival of natural law" and a bright representative of the Russian religious humanistic philosophy of law — all these titles are equally applicable to Novgorodtsev. They reflect different

1. The study was supported by the Basic Research Program at the National Research University Higher School of Economics (HSE University).

facets of the philosophical and legal heritage of the scholar who was notable for his broad and varied interests.

The philosophical and legal school led by Novgorodtsev was not a purely Russian phenomenon. It must be viewed within the context of the European legal thought in the first third of the 20th century, because "revived" natural law occurred not only in Russia, but in Germany and France as well. Novgorodtsev was deeply influenced by the ideas of prominent European thinkers in this field, mainly the professor at Berlin University, Rudolph Stammler, with his neo-Kantian interpretation of legal reality. The latter's concept of the ideal mission of law, its connection to morality, and his criticism of Marxism were especially inspiring to the Russian scholar. Novgorodtsev also paid considerable attention to the doctrines developed by theorists from the German historical school, G. Hugo, F. C. von Savigny and G. F. Puchta, coming to the conclusion that, in their later works, they admitted the importance of personal creativity and the juxtaposition of legal ideals with imperfect reality. He referred to the German legal scholar Rudolph von Jhering as well, who had come a long way from being fascinated with the ideas of the historical school, to criticizing them and becoming convinced of the power of human action and the need to fight for the law. Novgorodtsev was inspired by Jhering's conclusion that, in legal practice and when imposing sentences, one should be guided by the feeling of law, by the demands of a higher justice that is superior to the law (Novgorodtsev, 2010a: 127–128, 138, 162). Novgorodtsev makes multiple references to the theoretical heritage of Georg Jellinek, particularly his definition of law as the minimum of morality, the idea of the intrinsic value of the individual and the "positivization" of natural law in the process of law-making. Jellinek's knowledge of rich European legal tradition inspired Novgorodtsev to synthesize different lines of the philosophy of law and to seek common ground between them. This ground for him was the interpretation of law from the standpoint of the moral ideal.

Within the context of a retrospective analysis of European political and legal history in the first half of the 20th century, Novgorodtsev's doctrine appears to have been a predictor for the catastrophe of authoritarian regimes, which subjugated positive law to political will. It also appears to offer a means of resolving contradictions between politics and law. Novgorodtsev's search for a moral criterion to assess positive establishments seems relevant when we consider the turn of German jurists, Gustav Radbruch in particular, towards the idea of supra-legal law and to legal principles that are more authoritative than purely juridical prescriptions, as well as the jurisprudence of values which emerged after the collapse of the Nazi regime in post-war Germany.

It is noteworthy that Novgorodtsev had a major impact on European legal thought through his Russian followers: Ivan Ilyin, Boris Vysheslavtsev, Georges Gourvitch, Nikolay Alekseev, Sergey Hessen and Georgiy Florovsky. They immigrated to Western Europe after 1917 and brought many of the scholar's ideas to European soil, in particular his belief that liberal philosophy does not contradict the religious idealist tradition.

Novgorodtsev, who worked in the era of constitutional reforms, was not merely an armchair scholar. The lines of his academic and sociopolitical activities are closely inter-

twined. The creator of the philosophy of law of Russian liberalism was a prominent public figure: a member of the "Union for Liberation", one of the founders of the Constitutional Democratic Party and a member of its Central Committee, a deputy of the State Duma First Convocation, the author of bills on personal immunity, civil equality and freedom of assembly. The scholar's efforts when in emigration contributed to the establishment of the Russian Law Faculty at Charles University in Prague.

Although Novgorodtsev was born more than 150 years ago, his ideas are still the subject of intense debate. Philosophers, legal scholars, and historians are still trying to understand the phenomenon of Novgorodtsev and determine his place in the history of Russian legal thought. The subject of the greatest scholarly controversy is the continuity between his views in the earlier and later periods of his practice.

Novgorodtsev's last works primarily focused on religious and philosophical issues. Many of them were reconsidered and seem to have been presented in a different light, which prompts the conclusion that the philosopher shifted away from liberalism towards religious philosophy. Did the 1917 Revolution cause Novgorodtsev to give up his views and how far did he move away from the ideas of the philosophical and legal school, which he had designed and established at the turn of the century? There are different, often conflicting, views on this subject.

For instance, philosopher Andrzei Walicki, acknowledging that Novgorodtsev's idealization of Russian Orthodoxy during his exile was a new turn in the scholar's mindset, believes that this turn "did not cause any change in his view of the public ideal" and was merely a reflection of his views in the religious form (Walicki, 2012: 412–413). A similar observation was made by Kirill Soloviev (Soloviev, 2010: 18) who wrote that it was natural for the scholar who elevated legal norms to the level of absolute truth to take the position of an Orthodox philosopher. Meanwhile, according to the legal scholar Andrei Poliakov, during the period of revolutionary upheavals, Novgorodtsev "reconsidered his values" and adopted the position of Orthodox legal philosophy that seemed unable to accommodate the idea of natural law in its classic West-European interpretation (Poliakov, 2018: 75). Philosopher Viacheslav Zhukov describes Novgorodtsev's ideas from the later period of his life as "a conservative and religious utopia" (Zhukov, 2012: 68–77). This interpretation does not seem fair. Novgorodtsev's philosophy of law has regularly proven its validity in the 20th century and does not meet the criteria for a utopian theory as proposed by the philosopher himself. The scholar deemed utopian those theories that were built on rational and non-spiritual foundations; in his opinion, they were doomed to failure.

It is obvious that Novgorodtsev's ideas evolved in the direction of religious and humanistic philosophy, and exploring their origins reveals the connection between this philosophy and the secular liberal philosophy of law. Could the scholar work in both of these "fields", support the values of anthropocentrism and the primacy of human rights, while keeping to the conviction that the enhancement of the individual and the progress of the state are impossible without a synthesis of law and morality, attainable through church and religious communication?

We do not pretend to be pioneers in setting this research agenda. It was set, among others, by the historian of public thought Randall Poole in his paper on Evgeniy Trubetskoi. Evgeniy Trubetskoi was a prominent Russian liberal and a religious philosopher. Poole shows how Trubetskoi construed the Russian national identity from the standpoint of a synthesis of religious philosophy (metaphysics of unity), Kantian transcendental idealism and philosophical liberalism (Poole, 2016: 196).

Of particular relevance to the setting of the issue were the 1920s debates on the philosophical grounding of human rights. Their active participant, professor of law and history of Harvard University Samuel Moyn attributed the origin of the universal human rights doctrine to the catholic thinker Jacques Maritain (Maritain, 2011: 63–138). In Moyn's theory, Maritain is presented not only among the founders of the philosophical movement in France known as personalism, but also as the intellectual architect of the 1948 Universal Declaration of Human Rights and the original version of Christian (or "integral") humanism (Moyn, 2010, 2011, 2015). Randall Poole went further than Moyn. He substantiated the claim that, several decades before Maritain, the status of the human rights theory had been given to another influential tradition of Christian personalism. It was the Russian neo-idealism of the late 19th-early 20th centuries that combined personalism in its Orthodox Christian version with Kantian idealism and the doctrine of human dignity. The leading figure of this movement was Vladimir Soloviev, a religious philosopher, whose intellectual legacy, together with that of other idealists', was brought to interwar France by Nikolai Berdyaev and other prominent Russian emigre philosophers, contributing to the emergence of an intellectual environment for Maritain's teaching. Soloviev's Christian humanism and Maritain's integral humanism share the same roots (Poole, 2019). According to Poole, Pavel Novgorodtsev can also be regarded as an ideological progenitor of Maritain: his idealistic theory of liberalism and personal rights is based on the political theology of individualism (Poole, 2013: 160).

It is important to mention the Russian philosophers A. A. Kara-Murza and O. A. Zhukova, who pointed out the consistency between the values of liberalism and Christian humanism in the Russian political culture of the late 19th — early 20th centuries. Their heroes were also public figures, such as Ivan Aksakov, Mikhail Stakhovich, Vasilii Karaulov and Petr Struve. They presented a convincing intellectual project for the synthesis between liberal and Christian values, where the freedom of the individual was seen as the basis of law, while the law itself was based on the values of the Christian culture, both in the East and the West (Kara-Murza, Zhukova, 2011).

Laura Engelstein, an expert in Russian history, gave this discussion a more global turn by applying the categories of tradition and modernity, religious and secular foundations, often interpreted as the values of the East and West. She demonstrated that in late imperial Russia the educated public and bureaucratic elite viewed the mechanical juxtaposition of East and West as irrelevant. Despite the archaic nature of its legal tradition and political system, Russia has progressed in tandem with Europe. Its religious ethos evolved alongside the political and the legal, together with the emergence of the principles of human rights and the supremacy of law in legal culture (Engelstein, 2001).

The above-mentioned approaches enable us to conclude that the new liberalism and Christian humanism were not opposed to each other. The authors of the present paper aim to construe an integrated idea of the philosophical and legal views of Pavel Novgorodtsev and to overcome mythological statements, according to which the scholar's views in the final period of his life are conceptually different from his earlier views. To put an end to the protracted debates among specialists, the authors will reconstruct the scholar's views expressed during the period of over two decades of the 20th century. They will reveal the names of philosophers, legal scholars and historians, who had an impact on him, identify the journals in which the scholar had gotten his works published, describe the political, legal, and moral categories he pondered and show the social and historical grounding of his ideas. The logic of the paper is determined by the key stages of Pavel Novgorodtsev's life and work that shaped his legal ideas.

The authors achieve this goal by significantly expanding the source materials. The primary sources for the present research include not only P. Novgorodtsev's papers of different periods in which he formulated the ideas of the public ideal which are traditionally studied by historians. The authors also collected and analyzed private documents. The most significant among these are Novgorodtsev's hand-written letters to V. E. Grabar', a scholar in international law, dating 1890-1906, in which the former expounded upon his philosophical, legal, and moral views. The authors also relied on archival and published materials dealing with the thinker's activities in Russia and Prague.

Teachers and Colleagues from Novgorodtsev's Inner Circle

Pavel Novgorodtsev was greatly influenced by Vladimir Soloviev[2], the founding figure of Russian philosophical spiritualism. Soloviev's ideas about the crisis of legal consciousness and the need to overcome it, about the positive goals of law and state and the correlation between law and morality became guiding principles for Novgorodtsev. They constituted the philosophical outlook of new liberalism.

Novgorodtsev took the death of the prominent philosopher on June 31, 1900, as a bereavement. Attending Soloviev's funeral, he acknowledged his affinity to the deceased scholar "not only due to the appeal of his philosophy and talents", but out of personal liking as well[3]. Novgorodtsev also dedicated the work "The Idea of Law in the philosophy of V. S. Soloviev" to him. The paper formed the basis for his speech in memory of the philosopher during the ceremonial meeting of the Psychological Society on February 2, 1901. It was symbolic that this was Novgorodtsev's first public engagement. The scholar described the event in the following way: "On February 2 I was speaking in front of a large audience in Moscow for the first time. I should have got used to public speaking, as

2. Another person who was influenced by Soloviev's ideas was Evgenii Trubetskoi, mentioned above. The philosopher confessed that his philosophical and religious outlook "was full of Solovievian content and expressed in words close to Soloviev's". E. Trubetskoi was the author of a classic two-volume book on the philosopher, "The worldview of Vl. S. Soloviev". See: Poole, 2006: 195–240.

3. Otdel rukopisej Rossijskoj gosudarstvennoj biblioteki (OR RGB) f. 376 (Arhiv ya), K. 8 (P. I. Nogorodtsev, letters to V. E. Grabar), № 1 (1900), l. 18) (letter from P. I. Novgorodtsev to V. E. Grabar on August 9, 1900).

I did not feel the slightest anxiety before or after the speech… My speech at the Psychological Society was about Soloviev"[4].

In this memorial speech Novgorodtsev calls Soloviev "the most prominent defender of the idea of law among the philosophers in the past century". He argued that Soloviev saw the aim of law as the promotion of the moral principle among people. He presented the moral purpose of law as "suppressing evil inclinations…, fighting injustice and tyranny of those in power, ensuring equality and freedom". According to the philosopher, Soloviev's interpretation of the ideal content of law helped to revive the moral authority that had been lost by jurisprudence, and to affirm the value of law and the state within a religious outlook (Novgorodtsev, 1901: 1–2, 12).

Soloviev's definition of law as the minimum of "publicly organized morality" held an appeal for Novgorodtsev. Another thing that he found appealing was the fact that Soloviev's moral philosophy was not mere speculation. The thinker considered the idea of law as the basis not only for moral progress, but for the fair organization of legal institutions. Novgorodtsev gave a positive review of the transformation of Soloviev's ideas in the 1880-90s, when the latter made a statement that true Christianity should go public, and alongside saving individual souls, it should demand social reforms. Reflecting on the conditions of moral progress he believed it necessary not only to improve oneself as an individual but also to make every effort to change the social order. The scholar argued for the significance of common prosperity for human freedom and safety, for attaining morality and curbing despotism. In Soloviev's view, moral freedom must be guaranteed by law and by the state; without due guarantees on their part, it remains a fiction (Novgorodtsev, 1901: 7–9).

Novgorodtsev shared Soloviev's conviction that legal institutions were indispensable for moral progress. He quoted the philosopher's statement on the primacy of the moral principle, interpreted theologically: "All the human actions and relationships must ultimately be guided … by the moral principle that we worship in churches and that we live by in our everyday life, i.e., the principle of love, free consent and brotherly unity." According to Novgorodtsev, Soloviev found the features of this "true universal Christianity" in Dostoyevsky (Novgorodtsev, 1901: 10). Novgorodtsev himself highly appreciated Dostoyevsky's works.

Another thing that Novgorodtsev valued was Vl. Soloviev's respectful attitude towards the ideal of the rule of law as seen in Western Europe, expressed in the early 1890s during debates with Slavophiles. Meanwhile, Novgorodtsev pointed out that for Soloviev the rule of law was not the "ultimate manifestation of moral solidarity, but only a necessary step towards the ultimate form of theocratic communication" (Novgorodtsev, 1901: 8–9). This statement largely clarifies Novgorodtsev's own views.

Thus, Novgorodtsev was close to Soloviev in his ideas about the ethical nature of legal relations, which laid the foundations for the definition of positive law as "the guaranteed minimum of morality" (Medushevskii, 2010: 175). In the traditional methodological de-

4. OR RGB, f. 376, k. 8, № 2, l. 5 (letter from P. I. Novgorodtsev to V. E. Grabar on February 8, 1901).

bate about what is primary, ideas or facts, both thinkers argued for the primacy of ideas, sharing the view that what *should* exist must take precedence over what *does* exist and believing in the "future good", which was meant to transform the world.

Novgorodtsev considered Boris Chicherin to be his teacher and spiritual father. He acknowledged that Chicherin was the most outstanding representative of the older generation of Russian jurists, who was one of the first to speak about the crisis of legal consciousness, about lack of trust to the moral authority of the law, which was evident in the jurisprudence, and the benefits that the hard work of legal scholars brings solely for juridical practice. Instead of previous enthusiasm with sublime essence and ideal goals of law, it began to be regarded as a product of force and calculation (Chicherin, 1900: 24).

Like Vladimir Solovyev, Chicherin strove not only to grasp the abstract law of moral life, but also to study the practical forms of its implementation. Meanwhile, they had considerable disagreements, rooted in their differing views of human rights. As a representative of the "old", classical liberalism, Chicherin supported the traditional liberal doctrine based on the idea of "negative freedom", i.e., the sphere of individual self-expression free from state interference. In his work "The Justification of the Good" Vladimir Soloviev articulated his idea of the state's intervention in the realization of the right for dignified human existence. This view became the foundation for the doctrine of social and economic human rights, which was based on the positive function of the state and undermined the classical liberal doctrine of "negative freedom" (Walicki, 2012: 256).

Novgorodtsev shared Soloviev's views on this question, arguing that the fundamental legal value was the right to a decent life, formulated by Soloviev. Thus, as a prominent representative of neo-idealism in Russian philosophy, Novgorodtsev was a follower of Chicherin, and as a theorist of new liberalism he was influenced by the ideas of Soloviev.

In his personal correspondence with Vladimir Grabar' Novgorodtsev described himself as belonging to the "Chicherin School". He characterized this line of thought as something different from the "Russian sociological school" in legal science. In his opinion, the latter's most prominent representative was Sergey Muromtsev — a scholar of "brilliant, exquisite mind, whose philosophical potential was considerably higher than that of his associates from the 'positivist school'"[5]. The "Chicherin School", though, was not in an antagonistic relationship with the "Russian sociological school"; its language of communication was "poetic" rather than sociological.

Novgorodtsev had a personal acquaintance with Chicherin. The young scholar paid visits to the renowned jurist, met him at the homes of his colleagues during Chicherin's visits to Moscow in winter months, in particular, at the home of the historian Vladimir Gerie[6]. Novgorodtsev named his first son, born on January 14, 1904, shortly before Chicherin's death, Boris, after his teacher[7]. Evaluating Chicherin's role in developing philosophical thought in Russia, Novgorodtsev called him an outstanding philosopher-

5. OR RGB, f. 376, k. 7, № 16, l. 34. (letter from P. I. Novgorodtsev to V. E. Grabar on March 16, 1900).

6. OR RGB, f. 376, k. 7, № 16, l. 3 (letter from P. I. Novgorodtsev to V. E. Grabar on January 20, 1900), l. 15 (letter from P. I. Novgorodtsev to on February 4, 1900).

7. OR RGB, f. 376, k. 8, № 5, l. 1. (letter from P. I. Novgorodtsev to V. E. Grabar on January 22, 1904).

idealist who was one of the first in Russia to stand up against Anglo-French positivism and expose the flimsiness of its foundations. In describing Chicherin's success formula, his follower pointed not only to his exclusive talents and the depth of his scientific aspirations, but also to the moral enthusiasm that the scholar combined with his academic endeavors (Novgorodtsev, 2000: 575, 578).

In his 1901 speech in memory of Vladimir Soloviev Novgorodtsev awarded the title of the most outstanding Russian jurist to Lev Petrazhitskiy (Novgorodtsev, 1901: 23), who was the most consistent critic of juridical positivism and defender of the ethical function of law. Novgorodtsev credited him for recognizing a profound crisis of legal consciousness. He could also relate to Petrazhitskii's theory of legal policy. In this theory, Novgorodtsev discerned a revival of the original meaning of the category of "duty" in legal science. He saw Petrazhitskiy's doctrine as a crucial step in the development of natural law. A dogmatic analysis of the normative content of positive law, i.e., legal positivism, was seen by Novgorodtsev as the first step in the cognition of law. This was followed by the scrutiny of the psychological content of the idea of law, proposed by Petrazhitskiy. The final step was an ethical theory of law based on philosophical idealism, advanced by Novgorodtsev himself (Walicki, 2012: 357, 386).

While in St. Petersburg, Novgorodtsev and Petrazhitsky frequently dined together, they also shared their books. Meanwhile, he confessed to having a doubtful attitude towards Petrazhitsky[8]. It was accounted for by the traditional rivalry between the Moscow and Saint-Petersburg schools of legal philosophy to which the scholars belonged, as well as by the differences in their interpretations of the legacy of German idealism and understanding of law: Petrazhitskii understood it, for the most part, as a fact of individual and social psychology, while Novgorodtsev interpreted it from the positions of the metajuridical concept of justice.

Novgorodtsev's circle also included the Trubetskoi brothers: Sergei and Evgenii. Novgorodtsev attended the doctoral dispute of Sergei Trubetskoi in March 1900 and gave a rapt review[9]. The scholar hoped Trubetskoi would support him with his own doctoral thesis[10]. He also had some of his papers published in the journal "Issues in Philosophy and Psychology", edited by Trubetskoi. Novgorodtsev's speech at the meeting in memory of Vladimir Soloviev in 1901 was also made on Sergei Trubetskoi's request[11]. Novgorodtsev maintained a personal relationship with Evgenii Trubetskoi as well. He also paid close attention to the newspaper "Moscovskii Ezhenedel'nik" ("Moscow Weekly"), launched by Evgenii Trubetskoi in 1906, which focused on social and

8. OR RGB, f. 376, k. 7, № 16, l. 29 (letter from P. I. Novgorodtsev to V. E. Grabar on February 17, 1900), 47; 1902, k. 8, l. 32 (letter from P. I. Novgorodtsev to V. E. Grabar on March 16, 1900).

9. Sergei Trubetskoi's doctoral thesis was focused on the doctrine of Logos. After defending this thesis, he was awarded a doctorate and the position of an extraordinary professor at the Department of Philosophy of the Historical and Philosophical Faculty of Moscow University. According to Novgorodtsev, S. Trubetskoi's doctoral dispute "went brilliantly". Cited by: OR RGB, f. 376, k. 7, l. 37 (letter from P. I. Novgorodtsev to V. E. Grabar on March 30, 1900).

10. OR RGB, f. 376, k. 8, № 1, l. 3 (letter from P. I. Novgorodtsev to V. E. Grabar on June 4, 1900).

11. OR RGB, f. 376, k. 8, № 1, l. 1 (letter from P. I. Novgorodtsev to V. E. Grabar on January 12, 1901).

political issues and became a platform for philosophers of Russian spiritualism (E. Trubetskoi, S. Bulgakov, etc.), as well as liberal legal scholars and economists Alexander Alexeev, Pyotr Struve, etc. At the same time, Novgorodtsev was not in favor of the artistic circle that gathered around Evgeniy Trubetskoi and his friend Margarita Morozova, which included, among others, the poet Andrei Bely, the composer Alexander Skriabin, and the pianist Nikolai Metner. "Trubetskoi himself is pure, but there is a club behind him, which is useful, but not suitable for us[12]," the scholar wrote in January 1906.

Novgorodtsev had a close relationship with V. E. Grabar', an expert in international law. They corresponded for over 15 years, discussing all the events that happened in their professional and personal lives.

Novgorodtsev was also on friendly terms with Pyotr Struve. The scholar gave a very favorable review of Vladimir Hessen's works, considering them important for substantiating the ideas of the revival of natural law in Russia. In 1904 Novgorodtsev met Iosiph Pokrovskii, an expert in civil law, who was the Head of the Roman Law Department at St. Petersburg University. They had previously corresponded, and Novgorodtsev admitted that Pokrovskii was part of his circle (or "young faculty", as Novgorodtsev referred to his close circle in the winter of 1900-1901[13]).

To his "young faculty" Novgorodtsev assigned professors, experts in political economy, Alexander Manuilov and Ivan Ozerov, who lectured at Moscow University's Faculty of Law. He complained that among his colleagues and supporters there were no specialists in criminal law[14]. The most detailed description of Novgordtsev's circle of friends can be found in his diary entry from November 1903. Inviting his colleagues and friends to Kiev to celebrate Christmas, the scholar listed the people who belonged to his circle. He wrote to Vladimir Grabar', then a Professor at the Department of International Law at Yuriev University: "From your side we are waiting for you and Diakonov (Mikhail Diakonov was a historian of law, who taught at Yuriev University — *A. T., A. A.*), from Petersburg — Pokrovskii and Grevs (Ivan Grevs was a specialist in medieval and Roman law, who also taught at St. Petersburg University — *A. T., A. A.*), Kareev (Nikolay Kareev was a professor of European history — *A. T., A. A.*) and Lappo-Danilevskii (Alexander Lappo-Danilevskii belonged to the St.Petersburg school of Russian historians and methodologists of history — *A. T., A.* A.), from Moscow — besides me, there are Trubetskoi and Manuilov, in Kiev — Evgeniy Trubetskoi and Bulgakov (Sergey Bulgakov was then a professor at Kiev Polytechnical Institute and Assistant Professor at Saint Vladimir Kiev University — *A. T., A. A.*)[15].

12. OR RGB, f. 376, k. 8, № 6, l. 17 (letter from P. I. Novgorodtsev to V. E. Grabar on January 4, 1906).

13. OR RGB, f. 376, k. 8, № 5, l. 8 (letter from P. I. Novgorodtsev to V. E. Grabar on April 9, 1904).

14. OR RGB, f. 376, k. 7, № 16, l. 6 (letter from P. I. Novgorodtsev to V. E. Grabar on January 16, 1900); l. 19 (letter from P. I. Novgorodtsev to V. E. Grabar on February 10, 1900).

15. OR RGB, f. 376, k. 8, № 4, l. 6 (letter from P. I. Novgorodtsev to V. E. Grabar on November 20, 1903).

"You can't even imagine how burdened I am now with challenging and responsible work…"

The appeal to the moral ideal was the key idea of the philosophical volume "Problems of Idealism", which Novgorodtsev prepared in October 1902 in collaboration with Sergey Bulgakov, Evgeniy Trubetskoi, Nikolay Berdyaev, Bogdan Kistiakovskii and others. This work was a tipping point in Russian social philosophy. It gave rise to the religious and philosophical renaissance of the early 20th century in Russia. Novgorodtsev was the scientific editor of the publication.

Novgorodtsev, as he himself admitted, devoted "all his summer leisure" of 1902 to the "Problems of Idealism". On May 30, 1902, he moved to a country house, which was a two-hour journey from Moscow. In his own words, it was "a very pleasant location, where you could have a good time". Alexander Chouprov and Fyodor Kokoshkin also had their summer residences there. Novgorodtsev noted that he was held up in Moscow by the publication of the philosophical volume, which was completely his responsibility: "Our co-workers are Berdyaev, Bulgakov, Struve, both of the Trubetskoi brothers, Frank, Kistiakovskiy, Sergey Ol'denburg (on Renan), Lappo-Danilveskii (on sociologist Auguste Comte) and I. I am writing, I feel as if I were as Gelon[16]. At that moment I ceased being quiet…"[17]. In late August 1902 he reported that he was "burdened with challenging and responsible work of an urgent character"[18].

Novgorodtsev linked the publication of "Problems of Idealism" with the revival of public interest in idealist philosophy. In the introduction to the volume, he discusses "the eternal spiritual need" for moral consciousness, as well as the profound life journey towards moral regeneration. The problem of duty, formulated by the German philosopher I. Kant, was the central idea of the volume. Its key principle was the recognition of the unconditional value of the individual (Kolerov, 2018: 9–11).

In his essay "Moral Idealism in the Philosophy of Law" Novgorodtsev argued that legal scholars should considerably redesign the law. He described the condition of the current juridical science as a crisis, writing that "the critical spirit and high philosophical aspirations have been muted". He urged legal scholars to create a new juridical discipline with a broad system of scientific notions reflecting new existential bases and future ideals. These goals implied the revival of natural law with its *a priori* method, ideal aspirations, and recognition of the moral basis as an independent value (Novgorodtsev, 2018: 288–290).

This new line differed from the classical doctrine of natural law in its idea of changeability of the legal ideal. It used the formula of "natural law with a changeable content" introduced by the German neo-Kantian scholar Rudolph Stammler to substantiate the social (legal) ideal. Novgorodtsev saw Stammler's formula as the only possible solution

16. Gelon was a Greek tyrant of the 5th century B. C., a fierce and enterprising soldier, who was in command of cavalry and played a key role in a number of battles, including the conquest of Syracuse, as well as governing that city and turning it into a prosperous place.

17. OR RGB, f. 376, k. 8, № 3, l. 17 (letter from P. I. Novgorodtsev to V. E. Grabar on May 30, 1902).

18. OR RGB, f. 376, k. 8, № 3, l. 9 (letter from P. I. Novgorodtsev to V. E. Grabar on August 28, 1902).

to the problem of natural law. Novgorodtsev also referred to the young Russian scholar Vladimir Hessen, who called this school of thought "evolutionary natural law", and to Boris Chicherin, who was among the founders of philosophical and legal individualism (Novgorodtsev, 2018: 294, 330).

The Central Issue of Natural Law is the Law of the Future

The new school in jurisprudence emerged out of the need to improve the law. This need was identified due to the imperfections of existing laws and institutions and, in Novgorodtsev's words, "the anticipation of new forms of life". The need for a new approach was most evident in times of crisis, when the old forms of life proved their decrepitude, and society was impatient for a new order. The doctrine of natural law rested upon the practical demand for a more perfect legal reality, where natural law would be superior to positive law and interpreted as the ideal norm and goal. The natural legal constructions were, in the scholar's view, an intrinsic part of the human spirit and proof of humanity's higher vocation, as well as a sign of a healthy society: "A society that has ceased to create ideal construals would be a dead society; these construals regularly demonstrate that there is spirit in it, that there is a movement of moral feeling and consciousness (Novgorodtsev, 2018: 291–292)".

The idealist legal philosophy, promoted by Novgorodtsev, was supposed to reach the ambitious goal of creating the law of the future, understood as "the ideal and desirable legal reality, conforming to the moral law". Shaping the future was also the evaluation of the present from the standpoint of the category of moral duty (Novgorodtsev, 2018: 296).

The idea of moral foundations of the law is the cornerstone of Pavel Novgorotsev's work "Moral Idealism in Legal Philosophy", which became the manifesto of philosophical and legal idealism. For the scholar, morality was an "absolute value". He believed that legal norms must reflect the moral condition of the society: "the quality of these norms, their conformity to the moral basis predetermine whether the individual finds … a depressing lack of freedom or a happy opportunity to develop his potentialities". The scholar saw the goal of jurisprudence as "… setting moral requirements, prescribing ideal ways of development" to law, society, and the state. He called this approach to law a normative and ethical consideration, shaped by the ongoing revival of metaphysics and moral philosophy (Novgorodtsev, 2018: 292, 296, 330, 336).

Substantiating the ideology of the new doctrine, Novgorodtsev dissociated himself from the existing legal schools: historical, sociological, and positivist. He proved that the moral issue was "irresolvable in terms of historicism and was inaccessible to the positive method". Novgorodtsev denied a heuristic potential to the methodological apparatus of positive sociology, which did not exclude the variability of moral phenomena, but did not analyze the essence of morality, either. It appears quite natural that Novgorodtsev turned to the philosophy of Kant, which, as he said, "was the first to draw a line between existence and duty" and set "a dualism between these fields". Establishing a link between these fields was described by Novgorodtsev as the goal of modernity. To over-

come their dualism, he urged philosophers to turn to metaphysics, aimed at serving the ultimate good and realizing the moral law (Novgorodtsev, 2018: 297, 309, 317, 341–342). While Kant's idea of the dualism of existence and duty can be defined from the dialectical standpoint as a thesis, the desire to harmonize them proclaimed by Novgorodtsev must be recognized as a dialectical synthesis.

The correlation between law and morality was the subject of Novgorodtsev's earlier work, written in 1899. It defined law as a phenomenon inseparably connected with morality. Delimiting these notions (law, according to the scholar, always aims at exact definitions, while morality demands freedom), Novgorodtsev pointed out their flexibility, explaining that as society developed and the principle of moral freedom won over legal enforcement, they would come closer to each other. As the scholar said, the moral element was inherent in law, otherwise it mechanically transformed into a "product of force". Law must be supported by moral consciousness which gives it landmarks and brings it closer to ideal requirements (Novgorodtsev, 1899: 297, 129–132).

It is highly symbolic that in his paper written on the eve of the new century the scholar was concerned with the future. He supposed that as time passed the interaction between law and morality would intensify and they would form a unity in the long-term perspective. The legal norms would more adequately reflect the demands of moral consciousness, and inner sense would more frequently and more fully inspire the external implementation of law (Novgorodtsev, 1899: 297, 133–136). At the same time, Novgorodtsev believed that the boundary between law and morality would remain intact. He corrects Soloviev's view of law as a "minimum of moral requirements" and defines it as an enforced minimum of social requirements which are not exclusively of a moral character (Walicki, 2012: 378).

Novgorodtsev's Philosophical and Legal Substantiation of the Constitutionalist Ideology in Russia

In the autumn of 1900 Novgorodtsev got a book by Nikolay Berdyaev about Nikolay Mikhailovsky with a foreword by Pyotr Struve, where the recent leader of legal Marxism and champion of the objective approach to the social sciences moved on towards metaphysics. "What a gratifying fact!" was Novgorodtsev's reaction to Struve's methodological evolution. Struve's turn to metaphysics was accompanied by his references to Kant in the foreword, as well as his active support of the liberal doctrine of human rights (Walicki, 2012: 373–374). In the summer of 1900 Struve got in touch with a constitutionally inclined group of Zemstvo liberals and began to consolidate the forces opposing autocracy and sharing the need for political freedom (Pipes, 1970).

Among legal scholars' aspiration to political and legal renovation, renewal also manifested itself in the emergence of the school of the revival of natural law. The call for the renewal of the legal theory was voiced in Vladimir Hessen's article in the newspaper "Pravo" ("Law") on February 4, 1901. Timed to the start of the new century, this publication analyzed the progressive and destructive tendencies of juridical development in the

19th century. The legal activities in the late 19th century, Hessen believed, were not of a constructive character, as they did not contribute anything new to Russian life due to a lack of trust in abstract principles and an excessively pragmatic attitude to lawmaking. Improvements in the legal sphere were related to the development of the idealistic philosophy of law, capable of inspiring people with "high and beautiful ideals of law" (Hessen, 1901: 297–298). In his paper "Revival of natural law" published a year later Hessen admitted that the main symptom of the progressive development of law was the philosophical idealism in the works of Novgorodtsev, Petrazhitskii and Struve. Going back to the moral criterion of law, they symbolized progressive changes in political and public life (Hessen, 1902).

1904 saw the creation of "The Union of Liberation", a liberal organization which was the predecessor of the Party of People's Freedom, a prominent liberal party in pre-revolutionary Russia. Pavel Novgorodtsev became a council member of this organization and participated in its congress in July 1903 in Switzerland, together with Pyotr Struve, Nikolay Berdyaev, Semyon Frank, Ivan Petrunkevich, Bogdan Kistiakovsky and others. At the congress, it was decided that the newly established organization would fight for attaining political freedom. The political activities of the scholar were combined with his research into the theoretical and legal implementation of liberal reforms in Russia.

In 1904 Novgorodtsev was already a consistent supporter of the natural law revival doctrine. In the spring of 1904, he pointed out that he was preparing his own course on natural- law construction of the state. He also wrote articles about Immanuel Kant for the Moscow Psychological Society and the Petersburg Philosophical Society. Additionally, Novgorodtsev prepared a collection of papers for the new popular science journals "Nauchnoe slovo" ('Scientific Word") and "Voprosy zhizni"[19] ("Life Issues"). Active contributors of the former were Sergey Trubetskoi and Vasily Kliuchevskii, while the latter was favored by religious philosophers Sergiy Bulgakov and Nikolay Berdyaev.

Novgorodtsev acknowledged that 1904 had been a particularly productive year for scientific endeavour and that he had thoroughly enjoyed his work, despite there being more of the latter than he could cope with. In June 1904 the scholar confessed that he spent "most of the day among books writing a big paper, about 4-5 author's sheets, for "Pravo I gosudarstvo" ("Law and the State"). One of his immediate plans was the preparation of papers on the legal ideal and the relationship between the individual and society. The scholar meant to prepare them for the book "On the Doctrine of the legal ideal", which he was planning to complete soon[20].

In his paper "The Law and the State", published in the journal "Voprosy Filosofii i Psikhologii" ("Issues of Philosophy and Psychology") in the late 1904–early 1905 Novgorodtsev further substantiated his doctrine. This journal was of religious and philosophical character; it was edited by idealist philosopher Sergey Trubetskoi, among the contributors were Vladimir Soloviev, Boris Chicherin, Trubetskoi brothers, Vasily Rozanov, Nikolay Berdyaev, Gustav Shpet, Sergiy Bulgakov and others. The publication of

19. OR RGB, f. 376, k. 8, № 5, l. 8 (letter from P. I. Novgorodtsev to V. E. Grabar on April 9, 1904).

20. Op. cit., l. 11 (letter from P. I. Novgorodtsev to V. E. Grabar on July 29, 1904).

Novgorodtsev's paper in a journal of this kind was additional confirmation that the author belonged to the idealistic school, close to the religious thought.

Novgorodtsev's attention was primarily focused on the criticism of the formal juridical approach to the state. He argued that natural law was superior to the state and to positive law; it guided their activities. According to the scholar, the main factor of legal consciousness was the idea of the norm or the idea of duty; it was in this idea that law got its highest authorization. In order to conform to the main principles of contemporary legal consciousness, the theory of law must be developed in the direction of its natural bases. Novgorodtsev again argued that society had a demand for a developing law that would satisfy its needs. He suggested placing the idea of law above any particular legal system, substituting "that which has already been formed with what is newly forming", and seeing the history of law as a constant process of collisions and interactions between the idea of law and its temporary manifestations (Novgorodtsev, 1904a: 424, 442, 447–449).

It is noteworthy that in 1904 Novgorodtsev was not only a theorist, but also an advocate of constitutional changes, and sought to lay the theoretical foundations for them. He wrote about the relativity of formal legal constructions pertaining to the omnipotence of the state, which were acceptable in theory but collapsed as soon as they began to operate in social reality, where what is juridically possible does not coincide with what is morally possible and conflicts with it. Novgorodtsev argued that the restriction of state power by human rights, inviolable and inalienable, met the contemporary legal consciousness, which also demanded the same restrictions (Novgorodtsev, 1904: 405, 420).

Novgorodtsev's interpretation of human rights is natural, juridical and at the same time theological and philosophical. All human beings, the scholar believed, are granted inviolable rights that derive from human freedom and autonomy in relation to the state. He proposed that the idea of the rights of the individual should be at the foundation of social and legal orders. "We cannot provide a definition of the state without introducing the principle of freedom," Novgorodtsev wrote. The scholar saw the purpose and essence of law in defending personal freedom. He understood legal and political progress as the realization of the individual's natural rights, which did not require historical sanction due to being a direct manifestation of the moral law which was given before any experience. Such a neo-idealistic, Christian and personalist defense of human rights, as R. Poole points out, made Russian liberalism different from contemporary European liberalisms, for instance, British utilitarianism, which was more positivist in its philosophical foundations (Poole, 2019: 96).

Novgorodtsev paid considerable attention to the religious and theological interpretation of personal freedom. The individual, he believed, as well as the State, existed "through God's grace" and stood "before the trial of natural law". Personal consciousness was "the deepest root and source of the State's life". Human rights existed independently of their recognition by the State. The recognition of personal freedom by the state was, for the most part, negative, it meant no more than the official confirmation of a fact that demanded sanctioning with all the force of an unconditional phenomenon of life. The

self-restriction of the state had moral rather than juridical foundations and was dictated by the requirements of natural law (Novgorodtsev, 1904: 522, 527, 535, 537–538).

Novgorodtsev substantiated the primacy of human rights (natural law) over the law established by the State (positive law), and saw the realization of personal rights and freedoms as the ultimate goal in the development of both natural and positive law. This approach was an advance in the political and legal thought of Russian liberalism.

Along with the interpretation that denies the state's intrusion into the sphere of freedom, Novgorodtsev also justified an understanding of freedom linked to the active role of state power and its intervention into social relations. In formulating this principle the scholar referred to his contemporary new English liberalism, which aimed to provide people with a higher level of physical comfort and moral development. "Smith's and Bentham's old ideas about the all-healing effect of private interest, about the non-intervention of the state … have been left behind," Novgorodtsev argued. "The State is expected not only to eradicate juridical obstacles for the development of freedom, but also to provide a material opportunity for the best manifestation of freedom." The patronizing function of the state was primarily concerned with the provision of right to a decent life. Novgorodtsev characterized this right as granting every individual an opportunity of human existence and liberating them from the burdens of life that were killing them physically and morally. The recipients of this right were people suffering from economic dependence, lack of funds, and unfavorable conditions (Novgorodtsev, 2010b: 301, 303). Novgorodtsev interpreted it very broadly, including in its context the recognition of the right to care in case of disease, disability and old age, the right to work and to apply for a job, the right to a certain standard of living, etc. (Novgorodtsev, 1993: 129–131).

Like Soloviev, Novgorodtsev saw no contradiction between classic human rights and new social rights. At the same time, Novgorodtsev argued that the struggle for new social rights that implied an active position of the State (in the scholar's interpretation, the empowerment of social rights presupposed "state assistance rather than state intervention" (Novgorodtsev, 2010b: 313)), should not generate dangerous illusions about establishing heaven on earth (Walicki, 2012: 399).

Arguing against utopias and various kinds of utopian consciousness, Novgorodtsev took on the idea of the individual, recognizing his unconditional and sacred rights, for the sake of which it was expedient to make legal demands to restrict the common will and the majority principle. For him, the idea of personal freedom was a measure of democracy under any given political regime or state. He quoted the religious philosopher Nikolay Berdyaev, who warned that every time the individual and their freedom were made dependent on the will of men, every time the sovereignty of any human authority was recognized, the individual lost their absolute character, and their right to freedom fell under the burden of human passions and desires. Berdyaev wrote prophetically that "… the subjective will of proletariat, the tzar or any other human power can deprive the individual of the freedom of consciousness, the right to life and any other right." Novgorodtsev supplemented Berdyaev's idea with the recognition of personal freedom not only as the limit, but as the norm, the basis of the common will. Even if the ma-

jority decided to make a statement contradicting the idea of inalienable human rights, the scholar thought, the general agreement of all humans with it would not make it fair (Novgorodtsev, 2010b: 234–235).

The First Russian Revolution of 1905-1907 helped to deepen Novgorodtsev's views on legal consciousness as a factor of legal reform. In 1905 he joined the Constitutional and Democratic Party, and in the spring of 1906, he became a deputy of the State Duma, where he participated in drafting the laws on personal inviolability, civil equality and freedom of assemblies. Involvement in the legislative process strengthened the scholar's conviction that Russia needed profound legal reforms.

The Constitutional Democratic Party, to which Novgorodtsev belonged, set the goal of reforming the legal system of the Russian Empire. As he approached its practical implementation, Novgorodtsev found it challenging to introduce new statements into the body of the archaic Code of Law, many parts of which were imperfect and could not be transformed immediately. "When we laid down the new principles, that stemmed from the general requirements of the rule of law", he wrote, "we immediately discovered that these principles cut into our old laws, which were built on absolutely different principles. We had to build on the soil that was littered with old, widespread roots. As we prepared to abolish some laws, we immediately discovered that we had to think of abolishing or changing many other laws connected to them. And in an insuppressible impulse of thought that carried us forward, we embarked on the path … of hard work… the completion of which meant the reorganization of the entire Russian legal system" (Novgorodtsev, 1907: 2, 11).

The revolutionary period was characterized by a crisis of the public consciousness which manifested itself in a stark dissonance between positive law and the public ideal. In Novgorodtsev's opinion, the law could only be renewed if it abandoned its past; the scholar used the metaphor of "Saturn devouring his children". The resulting conflict gave rise to natural law, which urged people to change the existing order and proposed an ideal plan for social reorganization. In this logic, legal theories were presented as ideal projects of social reorganization (Novgorodtsev, 1904b: 36, 118).

Novgorodtsev saw the ideal of society in a rule of law characterized by legal equality, guarantees of individual rights and freedoms, opportunities for personal initiative, and the ability to resolve conflicts through reaching mutual consensus (Medushevskii, 2010: 177; Novgorodtsev, 2010b: 311). At the same time, for Novgorodtsev the category of the rule of law, as well as other political and legal categories, was based on the moral principle and interpreted as the "embodiment of the principles of justice". The rule of law was also presented as "the system of freedom", "an organization rising above the society and its members rather than absorbing them". The rule-of-law theory, as interpreted by the scholar, was incompatible with the "idea of a catastrophe that buries … the foundations of the state's legal structure for the sake of … creating of an absolutely new and previously unknown order, eliminating both the state and law". Demonstrating evolutionary thinking, Novgorodtsev believed that the "existing institutions and authorities should be transformed rather than abolished" (Novgorodtsev, 2010c: 718–719).

In his 1909 paper "Crisis of Modern Legal Consciousness" the scholar identified a crisis of the rule of law ideology based on classical liberalism. The paper was written within the paradigm of European new liberalism. According to the scholar, such foundations of the rule of law theory as public sovereignty and individualism were in acute crisis. The scholar saw the crisis of the idea of public sovereignty in the general disappointment with political democracy and representative government. The crisis of individualism manifested itself in the movement for positive freedom which was gaining momentum in the European society and in the tendency to extend the functions of the state in order to create mechanisms for the implementation of social rights (Walicki, 2012: 398).

What he was implying was a criticism of the classic ideas of rule of law because of the emerging need to extend its functions by regulating social life. Many theorists and advocates of the rule of law doubted its ability to resolve societal tensions, while its opponents spoke of its total unsuitability for ideal purposes. Admitting a crisis of individualism, the scholar still believed that the idea of the individual and their inalienable rights must be the basis of a morally acceptable social and legal order. For him, the symbol of progress in the historical development of the state was its ability to meet the needs of the individual (Novgorodtsev, 2010b: 235, 311, 365).

Thus, the rule of law was no longer seen by the scholar as "the pinnacle of history". He saw in it only a subordinate means, which was just an element among other moral forces. Just the same way as the law was incapable of embodying all moral requirements, the rule of law no longer seemed to Novgorodtsev to be the ideal of social development (Novgorodtsev, 2010b: 49). His previously voiced idea that the state is subordinate to superior norms and has to draw its guiding principles from them is the key point of the aforementioned 1909 paper.

In his report on the public ideal made at the Moscow Psychological Society in 1911, the scholar most explicitly expressed this idea. He argued that the public ideal required a public organization based on the principles of equality and freedom. This made sense, not by itself but because it created conditions for free personal development. It was the joint aspiration of people towards the absolute ideal that gave contemporary society its moral meaning (Zasedanie Psihologicheskogo obshchestva, 1911).

In his 1917 paper "On the Public Ideal" Novgorodtsev defined this ideal as the principle of free universalism, which reflected equality, human freedom, and their universality combined in the idea of free solidarity for all. Drawing on Soloviev's definition of the society as an internal free and universal agreement, Novgorodtsev supports the idea that when setting ideal goals of public progress, we should not disregard the dependency of the public principle on the unconditional idea of personality; it is in the fullness of individual consciousness that the philosopher saw the meaning and significance of public aspirations (Novgorodtsev, 2010c: 464).

As the influential theologian and philosopher Vasily Zen'kovskii later noted, the basis of this work was the religious and philosophical idea of the Absolute. According to him, Novgorodtsev's idea of "natural law" was based on ethics, which, in turn, was based on the principle of "absolute value", an absolute ideal (Zen'kovski, 1991: 127). Thus, on the

eve of the October Upheaval of 1917, Novgorodtsev, thinking in the spirit of the Christian Orthodox eschatological tradition, placed at the center of his public philosophy "the eternal ideal of the Good", the idea that each human being is endowed with the desire for a relentless pursuit of truth and is accompanied by the process of creation throughout his life and after his death. "No public, modern or ideal, will ever satisfy this desire", the scholar wrote prophetically, as if predicting the upcoming events. "In this relentless quest, man must be prepared not only for his ultimate hope, but also for his ultimate suffering" (Novgorodtsev, 2010c: 515).

Towards the Revival of the Sacred Values of Russian Life: the Transformation of Pavel Novgorodtsev's Views in the Post-Revolutionary Period and the Transfer of his Ideas to the European Intellectual Tradition

After the February Revolution of 1917 Pavel Novgorodtsev returned to Moscow University and actively worked in the Constitutional Democratic (Kadet Party). He belonged to those politicians who considered the Bolshevik revolution dangerous and argued for extreme measures to prevent socialists from coming to power. In August 1917 at a meeting of the Central Committee of the Kadet Party the scholar argued for the establishment of a military dictatorship, as he believed that only this could save the rule of the Provisional Government (Records of the Central Committee of the Constitutional and Democratic Party, 1915–1920, 1998). In July 1917, the scholar called the Central Committee to take decisive action, explaining their importance by the fact that "… the destruction is going on so rapidly that soon we will be in an abyss where can be no natural activity, for what remains in the abyss is moaning and shooting". Of particular note is the idea voiced by Novgorodtsev during his meeting with the Prime Minister Alexander Kerensky about Cadets joining the government: "When you have your own ideals, you always embed them into the framework of the state reality. The Church is not only a blessed realm and a community of the spirit. It also has its earthly forms. It is not the Church that asks from the State, but the State that gives. The duty of the State is to worship the sacredness of the souls of men". It was at this time that he remarked that the Cadets never demanded that the Church and State be separated (Kanischeva, 2012: 201–202).

The scholar refused to accept the October Revolution and engaged in an active struggle against the Bolsheviks. The majority of the professors of law from Moscow condemned new Bolshevik order in the autumn of 1917. On November 30, 1917, in response to the publication of the Decree on Court, which abolished the judicial institutions of the old regime, representatives of the Moscow Juridical Society, the All-Russia Union of Jurists, the Bar and the Prosecutor's Office assembled in Moscow University. Discussing current political events, the jurists characterized the Bolsheviks' behavior as a desire to destroy the whole order of life and as a "blow to the Russian legal consciousness" which, in legal terms, was throwing the country back into the days of "Russkaya Pravda" ("Russian Justice"). According to Novgorodtsev's colleague Iosiph Pokrovskii, the decree by the Soviet founder Vladimir Lenin abolishing the courts "basically destroyed any law

and justice, all public life was hung up in the air", and the country followed "the path of economic poverty, spiritual savagery and public disgrace". At that time, numerous lawyers saw their mission in defending the law and counteracting the new power as long as possible (Jurists' voice, 1917).

Pavel Novgorodtsev considered the Bolshevik ideology, which denied law and the legal tradition, to be utopian. In his words, the main function of the state was not to suppress, as the Bolsheviks claimed, but to harmonize social relations by establishing a unified and equal law to unite class, group, and personal interests. In his work "On Paths and Goals of the Russian intelligentsia", which was written for the papers collection "Iz glubiny" ("Out of the Depth"), illegally prepared in Moscow in the summer of 1918, he called the Bolshevik revolution the symbol of victory of the utopian consciousness. He saw its roots in the rationalist utopianism of the Russian intelligentsia, generated by socialist, anarchist and populist influences coming from Bakunin, Chernyshevskii, Lavrov and Mikhailovskii. These ideas led the Russian intelligentsia into non-religious apostasy from the idea of the state and to the collapse of intellectual consciousness. Novgorodtsev believed that the path followed by the intelligentsia at the time was disastrous and urged people to turn to the ideas of another school of thought popular among the Russian intellectuals and led by Chaadayev, Dostoevsky and Soloviev. It was rooted in the Old Russian culture and faith and Novgorodtsev felt it necessary to return to its values (Novgorodtsev, 2009: 823, 827, 835–839). The scholar started to deepen and develop these ideas in emigration.

When communicating with the Cadets who supported him in January 1918 the scholar admitted that the liberals had made mistakes: "We were too confident that the people could cope with the reform in its entirety. We were not conservative at all. It is not the people who are to blame, but we ourselves, because like youngsters we moved with the current. We thought that the idea of the state was enough, but we forgot about the required statecraft experience. And we did not have enough will for power. It should be taken into account that policy affairs do always have some ugly details" (Kanischeva, 2012: 215).

The year 1920 was fatal for many of Novgorodtsev's colleagues who supported the revivalist school of natural law. 1920 took the lives of Iosiph Pokrovskii, Vladimir Hessen and Evgeny Trubetskoi. Since December 1917 Pokrovskii had been working as a tenured professor at Moscow University and in July 1918 he wrote an article entitled "Perun's spell" for the anti-Bolshevik collection "Iz glubiny" ("Out of the Depths"), where he argued that the Russian revolutionary transition from monarchy to popular rule was turning into a dictatorship of the proletariat and the fall of the authority of law. He died of an asthma attack in April 1920. Vladimir Hessen was forced to leave Petersburg University after the revolution and went to the Social and Economic Faculty of the Ivano-Voznesensk Polytechnic Institute, where he worked only for a few months in 1919 and died of typhus in January 1920. Evgeny Trubetskoi was a member of anti-Bolshevik organizations in the south of Russia, but he also contracted typhus and died in January 1920.

Pavel Novgorodtsev survived the year 1920. Since the summer of 1918 he had been involved in the anti-Bolshevik movement in southern Russia. He tried to demonstrate to Western allies Bolshevism's global danger and the need for an immediate joint rejection of it. The scholar became an activist of the All-Russian National Center in the south of Russia, was the author of the main statements of this organization's program in 1919, as well as the draft declaration of the Volunteer Army on land and labor issues. The documents he wrote were full of the desire to overcome the tensions in the anti-Bolshevik movement. At the meetings of the Cadet Central Committee in Moscow in the spring and summer of 1918, he called for unity and saw the search for a "consolidated opinion" as an urgent task (Records of the Central Committee of the Constitutional and Democratic Party, 1915–1920, 1998: 445). The scholar contrasted the consolidation course aimed at "putting together power of the people" with the ideology of class division put into practice by the Bolsheviks. "There is no cadetism or democratism, but there is the national goal of unity," (Medushevskii, 2018: 406), the scholar argued.

Novgorodtsev taught at Simferopol University for a while, then immigrated first to Berlin and then, in 1921, to Prague. There he demonstrated his talents as a university manager, founding the Russian Law Institute (Faculty) at Charles University with the participation of the Czech government and becoming its first dean. In 1922, when filling in a personal questionnaire at the request of the Study Board of the Committee for the Education of Russian Students in the Czech Republic, the scholar wrote concisely and laconically: "Novgorodtsev Pavel Ivanovich… professor of public law, former professor at Moscow University and director of the Moscow Institute of Commerce. My special interest was legal philosophy and public law". When asked about his activity in Prague, the scholar answered, "I continue my research"[21].

At the Russian Faculty of Law he delivered lectures on the history of legal philosophy. Novgorodtsev also gave a series of lectures on the crisis of Westernism. In 1923 he published the paper "Democracy at the Crossroads" in the Russian philosophical collection "Sophia" and wrote a report of the same title in the Society for Jurisprudence. Since 1922 he was also the founder and the first chairman of Soloviev's Religious and Philosophical Society. In it, he made the report "The essence of the Russian Orthodox consciousness", which became the basis for a publication in the collection of Vasily Zen'kovskii, and gave a cycle of public lectures titled "Democracy and Dictatorship in General and in Socialist Literature"[22].

The papers written by the scholar during his emigration were full of religious and philosophical content and were prepared for the relevant collections. Among them were "The Orthodox Church and its Relation to the Spiritual Life of New Russia" (1922), published in "Russkaya Mysl'" ("Russian Thought") edited by Pyotr Struve, "The Essence of the Russian Orthodox Consciousness" (1923) for the collection of religious and philo-

21. State Archive of the Russian Federation (GARF), f. 5776 (Uchebnaya kollegiya pri Komitete po obespecheniyu obrazovaniya russkih studentov v Chekhoslovackoj respublike, Praga), op. 2, d. 110, l. 2.

22. GARF, f. 5776, op. 2, d. 110, l. 1.

sophical papers "Pravoslaviye i kul'tura" ("Orthodoxy and Culture") prepared by Vasily Zen'kovskii.

In his later works Novgorodtsev lamented that his political activities had indirectly contributed to the collapse of the old Russia. In his new circumstances, he continued his search for the public ideal and developed the conception that moral ideals should guide the development of law. However, the ideal type of the rule of law that he had construed in the pre-revolutionary period receded into the background, along with other questions of the external organization of human life.

In his works of the emigration period, Novgorodtsev associated the public ideal with the Russian Orthodox Church and religious communication. According to the ideologist of Eurasianism Georgy Florovskii, it was revealed by Novgorodtsev in the ecclesiastical conciliarity and in the new interpretation of social utopianism: he saw it "in the very spiritual organization of the European West" in connection with "the peculiarities of the Western religious realm". He understood that "… the crisis of public consciousness, of which he had spoken for so long, was essentially the crisis of the West" (Florovskii, 1924: 218).

The "spiritualization" of Novgorodtsev's philosophy of law in his last years was also noted by the young legal sociologist and his colleague at the Russian Faculty of Law, Georgii (Georges) Gurvitch, who later acquired influence in French sociology and jurisprudence. The ideological change evident in Novgorodtsev's position was seen by Gurvitch as a dialectical phase in the spiritual development of the scholar, who had initially belonged to the Westernist line of Russian public thought, but whose religious quest brought him closer to the Slavophiles. Gurvitch mentioned another change in Novgorodtsev's philosophy in the last months of his life, when the scholar turned again to the question of the "external construction of the human life" and tried to reconcile his new ideas with his earlier liberal Westernism. Gurvitch explained this synthesis of ideas by a profound spiritual crisis that the scholar experienced as a result of the revolution and emigration (Gurvich, 1924: 389–393).

The contemporary Russian legal philosopher Vyacheslav Zhukov has convincingly demonstrated that the concept of the absolute social ideal advanced by Novgorodtsev in his last years is derived from such fundamental categories of his legal philosophy as natural law and the moral law of the individual. The triad "moral law — natural rights — absolute ideal" is presented as the moral idea in its development, the evolution of the individual's ideas, which goes from the realization of one's own moral nature to the moral claims pertaining to the organization of society (Zhukov, 2012: 75).

In his paper "Restoration of Sacred Values", written in Prague in June 1923 (it was originally prepared as a speech in memory of Vladimir. D. Nabokov) Novgorodtsev set out his renewed idea of the Russian tradition in the philosophy of law. He saw its uniqueness in its decisive rejection of the foundations of the classic Western European legal tradition established in the 18th and 19th centuries and called for the development of the religious and ethical components of the public ideal (Novgorodtsev, 1926: 53–54).

Just like Iosiph Pokrovskii in his "Perun's spell" before him, Novgorodtsev admitted that his associates' "attempt to construct the rule of people" ended in failure and they "needed a new approach to life, a new attitude to reality". Novgorodtsev noted the destructive power of the revolution in shattering his contemporaries' faith in ideals. He placed the blame for it on the intelligentsia, whose worldview "was not shaped by Chicherin's state liberalism but was formed under the influence of Bakunin's populist anarchism". The Russian intelligentsia believed in the constructive power of revolution and in the creativity of the masses and thought that "it was only necessary to smash and destroy the old power…". The scholar saw this anarchical faith in many revolutionary activists: both in the leaders of the Provisional Government Grigory L'vov and Alexander Kerensky, and in the Bolsheviks.

Evaluating the events that happened in Russia after October 1917, Novgorodtsev argued that the revolutionary upheavals had enough potential to destroy Russia, but not to rebuild it. In his opinion, Russia's revival lay in the "reconstruction of sacred values", the first and foremost of them being the sacredness of the "people's soul", which was the key element of Russian statehood. "We must admit once and for all that the path of revolutionary 'conquests' has run its course, and that we must now take another path — that of gathering the Russian land and reconstructing the Russian state", the scholar argued. He linked the revival of sacred values with a dramatic change in political consciousness, a drastic change in the attitude towards the Motherland. The reconstruction of Russia could only be achieved in a united impulse of national unity, in a spirit of solidarity with higher principles, in a sense of responsibility for the whole (Novgorodtsev, 1926: 62–64).

Novgorodtsev thus called for the reconstruction of religious and national forces. The pathos of nationalism for him meant that "Russia and the Russian culture are superior to parties and political dogmas". The thinker noted that among young people "there was an unbeatable growth of national consciousness, strengthening natural attraction to their own, native things. A passionate love for their country and the Motherland was emerging. Long-forgotten words and feelings were in use again" (Novgorodtsev, 1926: 65). The religious revival was seen as a revival of the orthodox consciousness, based on the idea that each individual has a shared moral and religious responsibility for everyone else and vice versa. At the core of this conviction was the idea of a joint and conciliar salvation by the force of everybody's feat of faith, prayer, and love (Novgorodtsev, 1923: 11).

By the reconstruction of the sacred values the scholar understood not the revival of the external forms and ways of life, but, first and foremost, the revival of souls, a religious and moral renewal. "To create new Russia, we need new spiritual powers, we need souls that aspire to new light," wrote the philosopher. His search for support in the religious enlightenment of spirit was inspired by his conviction that the greatest crisis of legal consciousness is a global crisis, a "crisis of lack of faith, a crisis of culture divorced from religion, crisis of the State which rejects any connection with the Church, and a crisis of human law that renounces kinship with divine law" (Novgorodtsev, 1926: 67). Novgorodtsev predicted the difficulties of the religious revival in Soviet Russia, which was an authoritarian state.

Novgorodtsev pointed out that the Russian people would have to live and act "in a country that had been destroyed and thrown back several centuries", amidst not only the greatest material destruction, but also the "terrible collapse of cultural, public and economic foundations". The thinker concluded that "in the years ahead, the Russian people will have to make heroic and selfless efforts", as sowers of "the reasonable, kind and eternal" will have to act under terrible and primitive conditions, because the revolution will leave major devastations behind it, not just in the external environment, but in human souls (Novgorodtsev, 1926: 70–71).

Analyzing the experience of the Russian Revolution, Novgorodtsev concluded that legal systems are not autonomous and are powerless without moral upbringing. He argued that, on its own, morality does not possess autonomy and is not the highest principle. Once again, just as at the turn of the century, Novgorodtsev speaks of the need for developing legal consciousness: "It is not the political parties that will save Russia, it will be revived by the people's spirit that will aspire to the light of eternal sacred values!" "That is why", concludes Novgorodtsev, "we are now replacing autonomous morality with theonomic morality, and replacing democracy, the rule of the people, with hagiocracy, the power of the sacred. It is not some universal forms that will save us, but the blessed enlightenment of the souls" (Novgorodtsev, 1926: 69, 71).

The goal of "restoring sacred values" shaped Novgorodtsev's activities as dean of the Russian Law Faculty of Charles University in 1922–1924. In his speech at the opening of the faculty he said that it was a "bet on the future of Russia". Referring to the national character of Russian science, Novgorodtsev explained that Russian law students studying abroad would be able to broaden and enrich their knowledge base and put it at the service of their fatherland (Studencheskiye gody, 1922: 23–25). This mission determined the scholar's public activity in Prague, where he participated in the work of multiple scientific and pedagogical organizations of Russian immigrants. Trying to preserve the basics of Russian culture, Novgorodtsev took on the positions of creator and chair of Vladimir Soloviev Religious and Philosophical Society at the Russian Law Faculty, as well as member of the Board of the Union of Russian Academic Organizations Abroad. The scholar believed that the strongest spiritual bond between people raised within the context of Russian culture is a devotion to that culture.

* * *

To sum up, Pavel Novgorodtsev, the leader of the idealist line in legal philosophy, consistently supported the idea that moral ideals should guide the development of law. As the founder of the school of natural law revival in Russia, Novgorodtsev substantiated the ethical doctrine of law, which resonated with the ideas of the German legal scholars Georg Jellinek, Rudolph Stammler, Gustav Radbruch and others.

A prominent scholar and a political activist, Novgorodtsev skillfully brought his ideas into the context of contemporary law-making. In papers published at the time of the First Russian Revolution he sought to demonstrate the need for legal reforms aimed at ensur-

ing the natural and social rights of the individual. He drafted bills on personal inviolability, civil equality and freedom of assembly. Between 1907 and 1917, under the influence of the evolution of liberalism in European countries, the scholar advanced the ideas of the new liberal philosophy concerning the importance of expanding the functions of the state in controlling social life.

The scholar was one of the founders of the religious and philosophical movement in Russia at the beginning of the 20th century. He was one of the first Russian legal scholars to note the importance of Vladimir Soloviev's moral and legal ideas for the Russian legal tradition, and the editor of the 1902 iconic collection "Issues of Idealism". Under the influence of the Russian revolutions Novgorodtsev's philosophical views became increasingly religious. He turned to the criticism of utopian theories that subjugated the individual to society and equated law with power, admitting the relativity of the liberal ideal of the rule of law and searching for the absolute.

After 1917 Novgorodtsev became a full-fledged religious philosopher. His turn to Christianity as the source of justice and ethical norms was another attempt to search for the public ideal. It was the search for the ideal form in the internal organization of the society, in the individual, which was the only opportunity in conditions when any participation in transforming the external forms of life was no longer accessible to him. Novgorodtsev relied on the national spiritual tradition and advocated the preservation of its principles in the European cultural environment. This construction, in our opinion, did not contradict the idea of the rule of law, which was the core of the scholar's research in the previous period; the rule of law was the external form of social organization, but it was also necessary to find its internal foundations.

As the creator of the philosophy of law of "new liberalism" Novgorodtsev established the ethical doctrine which did not contradict liberal philosophy. In his creative work, new liberalism and Christian humanism relied on each other and laid the foundations for each other. At the heart of both approaches was the justification of individual autonomy in relation to the state, the idea of the ideal social order as the realization of the principles of legal freedom and justice, the view of law as the ultimate social and ethical value.

References

Chicherin B. (1900) *Filosofiya prava* [Philosophy of law], Moscow: Tipo-litografija Tovarishhestva I. N. Kushnerev i K°.

Engelstein L. (2001) Holy Russia in modern times: an essay on Orthodoxy and cultural change. *Past & Present*, no 173, pp. 129–156.

Florovskii G. (1924) Novgorodtsev kak filosof prava [Novgorodtsev as a philosopher of law]. *Modern notes*, vol. XX, pp. 218–224.

Gurvich G. (1924) Professor P. I. Novgorodtsev kak filosof prava [Professor P. I. Novgorodtsev as a philosopher of law]. *Modern notes*, vol. XX, pp. 389–393.

Hessen V. (1901) Glavnye momenty juridicheskogo razvitiya Rossii v XIX veke [The main moments of legal development of russia in the 19th century]. *Law,* no 6, pp. 282–298.

Hessen V. (1902) *Vozrozhdeniye estestvennogo prava* [The revival of natural law], St.-Petersburg: Tipografija Spb. o-va E. Evdokimov.

Jurists' voice (1917) *Russkiye vedomosti,* no 253.

Kanischeva N. (ed.) (2012) *Nasledie Ariadny Vladimirovny Tyrkovoj: Dnevniki. Pis'ma* [Heritage of Ariadna Vladimirovna Tyrkova: Diaries, Letters], Moscow: Rosspen.

Kara-Murza A., Zhukova O. (2011) *Svoboda i vera. Khristianskii liberalizm v rossiiskoi politicheskoi culture* [Freedom and faith. Christian liberalism in Russian political culture], Moscow: IF RAN.

Kolerov M. (ed.) (2018) *Problemy idealizma (1902)* [Problems of idealism (1902)], Moscow: Modest Kolerov.

Maritain J. (2011) *Christianity and democracy: the rights of man and the natural law* (trans. D. C. Anson, intro. D. A. Gallagher), San Francisco: Ignatius Press.

Medushevskii A. (2010) *Dialog so vremenem. Rossiiskiye constitutsionalisty kontsa 19 — nachala 20 veka* [Dialogue with time. Russian constitutionalists of the late 19th — early 20th century], Moscow: Novy khronograf.

Medushevskii A. (2018) Pravo i revolutsiya v politicheskoi filosofii russkogo konstitutsionalizma [Law and revolution in the political philosophy of Russian constitutionalism]. *Filosofiya prava: P. I. Novgorodtsev, L. I. Petrazhitskii, B. A. Kistiakovskii* [Philosophy of law: P. I. Novgorodtsev, L. I. Petrazhitskii, B. A. Kistiakovskii] (ed. E. Pribytkova), Moscow: Rosspen, pp. 394–427.

Minutes of the Central Committee of the Constitutional and Democratic Party, 1915–1920, vol. 3 (Moscow, 1998).

Moyn S (2011) Personalism, community, and the origin of human rights. *Human rights in the twentieth Century* (ed. S.-L. Hoffman), Cambridge: Cambridge University Press, pp. 85–106.

Moyn S (2015) *Christian human rights*, Philadelphia: University of Pennsylvania Press.

Moyn S. (2010) *The last utopia: human rights in history*, Cambridge: Harvard University Press.

Novgorodtsev P. (1899) Pravo i nravstvennost' [Law and Morality] *Sbornik po obschestvenno-yuridicheskim naukam* [Collection on social and legal sciences] (ed. Yu. Gambarov), S.-Petersburg: O. N. Popova's Publishing House, pp. 113–136.

Novgorodtsev P. (1901) *Ideya prava v filosofii Vl.S. Solovieva. Rech, proiznesennaya na torzhestvennom zasedanii psikhologicheskogo obschestva v pamiat' Vl.S. Solovieva 2-go fevralia 1901* [The idea of law in the philosophy of Vl. S. Solovyov. Speech delivered at the ceremonial meeting of the Psychological Society in memory of Vl. S. Solovyov on February 2, 1901], Moscow: Tipo-litografija Tovarishhestva I. N. Kushnerev i K°.

Novgorodtsev P. (1904a) Gosudarstvo i pravo [Law and state] *Questions of philosophy and psychology*, vol. XV, no 74 (IV), pp. 397 — 450.; vol. XV, no 75 (V), pp. 507–538.

Novgorodtsev P. (1904b) *Iz lektsii po istorii filosofii prava* [From a lecture on the history of philosophy of law], Moscow: Tipo-litografiya N. A. YAshkina.

Novgorodtsev P. (1907) Zakonodatel'naya deyatel'nost' Gosudarstvennoi dumy [Legislative activity of the State Duma] *Pervaya Gosudarstvennaya duma* [First State Duma], S.-Petersburg: Izdanie A. A. Muhanova i V. D. Nabokova, pp. 9–30.

Novgorodtsev P. (1923) Suschestvo russkogo pravoslavnogo soznaniya [Essence of Russian Orthodox consciousness]. *Pravoslaviye i kul'tura: Sbornik religiozno-filosofskikh statei* [Orthodoxy and culture: a collection of religious and philosophical articles] (ed. V. Zen'kovskii), Berlin: Russkaya kniga, pp. 7–23.

Novgorodtsev P. (1926) Vosstanovleniye sviatyn' [Restoration of shrines]. *The way*, no 4, pp. 54–72.

Novgorodtsev P. (1993) Pravo na dostoynoye chelovecheskoye suschestvovaniye [Right to a decent human existence]. *Social sciences and modernity*, no 5, pp. 129–138.

Novgorodtsev P. (2000) B. N. Chicherin [B. N. Chicherin]. *O svobode: Antologiya mirovoi liberalnoi mysli (1 polovina XX veka)* [On freedom: An anthology of world liberal thought (I half of the 20th century)], Moscow: Mysl', pp. 575 — 579.

Novgorodtsev P. (2009) O putjah i zadachah russkoj intelligencii [On the paths and tasks of the Russian intelligentsia]. *Manifest russkogo idealizma* [Manifesto of Russian Idealism], Moscow: Astrel', pp. 823–839.

Novgorodtsev P. (2010a) Istoricheskaya shkola yuristov. Yeyo proiskhozhdeniye i sud'ba [The historical school of lawyers. Its origin and fate]. *Nemeckaja istoricheskaja shkola prava* [German historical school of jurisprudence], Chelyabinsk: Socium, pp. 1-225.

Novgorodtsev P. (2010b) Vvedeniye v filosofiyu prava. Krizis sovremennogo pravosoznaniya [Introduction to the Philosophy of Law. Crisis of modern legal consciousness] *Novgorodtsev P. I. Izbrannye Trudy* [Novgorodtsev P. I. Selected Papers] (ed. K. Soloviev), Moscow: Rosspen, pp. 33–366.

Novgorodtsev P. (2010c) Ob obschestvennom ideale [On the social ideal] *Novgorodtsev P. I. Izbrannye Trudy* [Novgorodtsev P. I. Selected Papers] (ed. K. Soloviev), Moscow: Rosspen, pp. 367–898.

Novgorodtsev P. (2018) Nravstvennyi idealism v filosofii prava [Moral idealism in the philosophy of law]. *Problemy idealizma (1902)* [Problems of idealism (1902)], Moscow: Modest Kolerov, pp. 275–342.

Pipes R. (1970) *Struve. Liberal on the left, 1870–1905*, Cambridge, Massachusetts: Harvard University Press.

Poliakov A. (2018) P. I. Novgorodtsev i ideya "vozrozhdennogo yestestvennogo prava" kak kommunikativnaya problema [P. I. Novgorodtsev and the idea of "revived natural law" as a communicative problem]. *Filosofiya prava: P. I. Novgorodtsev, L. I. Petrazhitskii, B. A. Kistiakovskii* [Philosophy of law: P. I. Novgorodtsev, L. I. Petrazhitskii, B. A. Kistiakovskii] (ed. E. Pribytkova), Moscow: Rosspen, pp. 51–77.

Poole R. A. (2006) Religion, war, and revolution: E. N. Trubetskoi's liberal construction of Russian national identity, 1912–20. *Kritika: Explorations in Russian and Eurasian History*, vol. 7, no 2, pp. 195–240.

Poole R. A. (2013) Russian political theology in an age of revolution. *Landmarks revisited: The Vekhi Symposium 100 years on* (eds. R. Aizlewood and R. Coates), Brighton: Academic Studies Press, pp. 146–169.

Poole R. A. (2019) Integral humanisms: Jacques Maritain, Vladimir Soloviev, and the history of human rights. *Vestnik of Saint Petersburg University. Philosophy and Conflict Studies*, vol. 35, issue 1, pp. 92–106.

Soloviev K. (2010) Pavel Ivanovich Novgorodtsev. *Novgorodtsev P. I. Izbrannye Trudy* [Novgorodtsev P. I. Selected Papers] (ed. K. Soloviev), Moscow: Rosspen, pp. 5-32.

Studencheskiye gody (1922) *Ezhemesiachnyi zhurnal,* no 1, pp. 23–25.

Walicki A. (2012) *Filosofiya prava russkogo liberalisma* [Legal philosophies of Russian liberalism] (ed. S. Chizhkov, trans. O. Ovchinnikova, O. Pazuhina, S. Chizhkov, N. Chistyakova) (Moscow: Mysl', 2012), 412–13.), Moscow: Mysl'.

Zasedanie Psihologicheskogo obshchestva (1911) *Russkiye Vedomosti*, no 291.

Zen'kovskii V. (1991) I*storiya russkoi filosofii* [History of Russian philosophy], Leningrad: EGO, vol. II, part 2.

Zhukov V. (2012) Filosofiya prava P. I. Novgorodtseva kak konservativno-religioznaya utopiya [The philosophy of law by P. I. Novgorodtsev as a conservative-religious utopia]. *State and law*, no 2, pp. 68–77.

Философия права Павла Новгородцева: «новый либерализм» *vs* христианский гуманизм[23]

Анастасия Туманова
Доктор исторических и юридических наук, профессор, Школа исторических наук Факультета гуманитарных наук,
Национальный исследовательский университет «Высшая школа экономики»
Адрес: ул. Мясницкая, д. 20, Москва,101000 Российская Федерация
E-mail: atumanova@hse.ru,

Александр Сафонов
Доктор юридических наук, профессор, факультет права,
Национальный исследовательский университет «Высшая школа экономики»
Адрес: ул. Мясницкая, д. 20, Москва,101000 Российская Федерация
E-mail: asafonov@hse.ru

Авторы статьи характеризуют вклад философа Павла Новгородцева в становление российской философии права, в развитие моральной и этической традиции права, являвшейся философским обоснованием доктрины «нового либерализма» и прав человека. Эта традиция была авторитетной как в позднеимперской России, так и в Европе, куда Новгородцев эмигрировал в 1920 г. и где он основал юридический институт (факультет) при Карловом университете (Чехословацкая республика).

Авторы впервые предприняли попытку рассмотреть философию права Новгородцева как целостную этическую концепцию с учетом ее развития в завершающий период его жизни. Статья является одной из первых в академической литературе аргументаций идеи, что воззрения Новгородцева необходимо интерпретировать в контексте как либерально-философского, так и религиозно-философского подходов к правопониманию, и оба этих

23. Исследование выполнено в рамках Программы фундаментальных исследований в НИУ ВШЭ в 2024 г.

подхода взаимодополняют друг друга, сыграв существенную роль в истории политико-правовой мысли России и Европы XX столетия. Вклад Новгородцева в интеллектуальную мысль России и Европы заключался также в обосновании содержания социальных утопий, в частности, большевизма и коммунизма, характеризовавшихся отказом от права и правовой традиции.

Исследование базируется на широком круге источников, включающих труды П. Новгородцева и документы личного происхождения, в частности, его переписку.

Ключевые слова: русская интеллектуальная история, философия права, религиозная философия

doi: 10.17323/1728-192x-2024-4-249-258

«The Freedom to Be Free»: Hannah Arendt on «Salvation» of the World of Politics

Timofey Dmitriev
Associate Professor, Faculty of Humanities, HSE University
Address: Staraya Basmannaya str., 21/4c1, Moscow, 105066 Russian Federation
E-mail: tdmitriev@hse.ru

Publications of recent years and decades demonstrate that, in addition to writing monographs and publishing collections of articles, Hannah Arendt (1906-1975), throughout her academic career in the United States, actively used the lecture genre and non-fiction articles to clarify and popularize the ideas of her political philosophy and her approach to the problems of our times. The publication of these lectures provides us with a fresh perspective on important aspects of her ideas regarding the political world and the role of political freedom, as reflected in the words and actions of humans in their common world. The focus of this paper is on Arendt's lectures from the 1960s, which were dedicated to clarifying her two key works from the second half of the 1950s and the early 1960 — "The Human Condition (1958)" and "On Revolution (1963)" — as well as interpreting the central concepts of her political philosophy: Action, Freedom, Politics and Revolution. The article pays special attention to Arendt's interpretation of human freedom as a gift and a miracle, which is proposed in these lectures. Freedom's negative contribution to the world of human actions and words is to interrupt the automaticity of inertial social reproduction, while its positive contribution is to "save" the common political world for future generations.

Keywords: Hanna Arendt, political philosophy, action, freedom, world of politics, revolution, salvation, secular theology of redemption

In "The Life of The Mind", a book that Hannah Arendt wrote in her final years, she compared the work of human mind to the web of Penelope, a mythical character, who was the wife of the legendary Odysseus. "The business of thinking", she said, "is like Penelope's web; it undoes every morning what it has finished the night before" (Arendt, 1978: 88). These ideas are based on the assumption that the philosophical journey of the mind does not lead to any final conclusions that are not open to further reconsideration, but rather consists in an ongoing dialogue that the philosopher engages in not only with others, but primarily with himself throughout his lifetime, satisfying his need for thought through the very act of thinking. This principle is perfectly in line with Arendt's own ideas, as has been repeatedly emphasized by scholars of her political philosophy[1]. This trait of her thought is well illustrated by her lectures and articles from the 1960s, recently published by Penguin as a separate volume (Arendt, 2020) in the "Great Ideas" series[2].

1. Bearing in mind this precise feature of Arendt's thinking as a philosopher, Claude Lefort speaks even of the "requirement to think, which lies at the heart of her work" (Lefort, 1986: 65).

2. In addition to Arendt, the thinkers of the 20th century whose works were honored with publication in this series include Sigmund Freud, George Orwell, Thorstein Veblen, Albert Camus, Leon Trotsky, Walter Benjamin, Frantz Fanon, Michel Foucault, William James, W. E. B. Du Bois, Theodor Herzl, Vladimir Lenin, Peter Kropotkin, Simone de Beauvoir, Simone Weil, Martin Luther King, and Peter Singer.

These lectures and articles, although delivered and published at different times, were chosen in such a way that Arendt's thoughts unfold sequentially in three interrelated stages. The problems raised in the previous texts smoothly turn into those of the subsequent ones, creating an excellent impression of a coherent presentation. The first text provides an interpretation of Arendt's understanding of the conditions of human existence, as presented in her work "The Human Condition" (Arendt, 1998 [1958]). The second lecture reveals her understanding of politics as the highest form of human being-in-the-world, the source of which she sees in humans' ability of beginning something new and free from the pressure of vital necessity. In turn, Arendt sees revolutions — these fundamental events of the modern era, aimed at constituting a new public space for political freedom (*constitutio libertatis*) — to be the ultimate expression of such a new beginning in the political life of contemporary world[3]. The third text in this collection focuses on the background and importance of the great revolutions of the modern era. Thus, the drama of the political aspect of human existence, as discussed in Arendt's lectures and articles, takes place within a theoretical framework defined by her three main philosophical concepts: action, political freedom, and revolution.

Political thought, true to its purpose, for Arendt is always a test of events. She believed that, «the significance of a historical period shows itself only in the few events that illuminate it» (Arendt, 1998 [1958]: 42). Only by hermeneutically thinking and interpreting such fateful events, if you will, Events with a capital E, that define the essence of the modern era, can one truly understand its nature. Thoughts that evade comprehension of fateful events are objectively empty, and events that are not affected by thought comprehending their significance are meaningless in their bare factuality. "My assumption is", Arendt wrote in the preface to her work «Between Past and Future», "that thought itself arises out of incidents of living experience and must remain bound to them as the only guideposts by which to take its bearings" (Arendt, 2006 [1954]: 14). Arendt remained faithful to this idea that the philosopher's political thought should be substantially linked to the fateful events of our time, not only in her written works, but also in her public speeches, reports, and lectures.

At the same time, it's important to remember that the lecture genre, as the genre of non-fiction articles which requires presenting the most important information in a limited time frame and printed space and making it understandable to an unprepared audience, demands a lot from those who choose to share their thoughts publicly. In this regard, we, Arendt's contemporary readers, are very fortunate, because she was able to explain complex and structured ideas that were initially understandable only to academics and experts, in simple and clear language. This is exactly the case with her first text "Labor, Work, Action" (Arendt, 2020 [1964]: 1 — 32), published in this collection, which is dedicated to presenting the main outlines of her understanding of the human condition. In this text, Arendt comprehensively examines three basic forms of human practice: Labor, Work or Production, and Action. She explores their specifics, the unique features

3. Arendt addressed this issue a separate work "On Revolution" (Arendt, 1990 [1963]).

of their historical evolution, their interrelation with one another, and their impact on the political space.

To recap, in her philosophy, Arendt divides all human activity into two basic categories — theoretical (*vita contemplativa*) and active (*vita activa*) (Arendt, 1998 [1958]: 7 — 17). The latter, in turn, is further divided into three primary types — labour, production/manufacture (work), and action. If, in the course of their labour, humans are entirely dependent on the natural processes of life, compelled to reproduce their physical and mental existence and fulfill their basic requirements, then through the second form of practical activity — work — they «render themselves the masters and possessors of nature» (as Descartes put it) and shape the world of objects around them. However, it is only through the third form of human activity — action and speech — that an individual engages with others and the world, serving as the backdrop for this engagement, and can realize their freedom. By making themselves known to others through speech, humans thereby make their presence in the world public. This is an elementary action from Arendt's perspective.

Politics belongs precisely to this last, highest type of human activity. It is the deliberation and solution of common problems through speech, which involves the active participation of citizens in socially significant processes of verbal communication[4]. That is why, according to Arendt (Arendt, 1990 [1963]; Arendt, 1998 [1958]; Arendt, 2006), the normative model of a political community in her works is the ancient *polis* and the Roman *civitas*. Both represent a political community of free and equal individuals united by a shared commitment to a particular political way of life and a collective willingness to participate in solving common challenges.

This notion of interdependence between freedom and politics stands in contradiction to the social theories of the modern age. With the growing world-alienation of man in the modern age political action as a form of practice recedes into the background[5]. This is why politics is often mistakenly associated either with work, or with labor. Work refers to the manufacture of durable consumer goods, while labour refers to activities aimed at meeting immediate human needs. In this case, politics can be reduced to either bare administration, where speech-mediated interaction between people is likened to the management of things, or to a specific type of politics that, with the support of Michel Foucault and his followers, has been called "biopolitics"[6]. The result of these trends in the development of political practices and institutions, as well as political theory, in the

4.Arendt defines as "political" those matters of common public interest for which there is no certain technical solution and which therefore are a suitable subject for public debate. From her point of view, «Public debate can only deal with things which — if we want to put it negatively — we cannot figure out with certainty. Otherwise, if we can figure it out with certainty, why do we all need to get together?» (Arendt, 1979: 317).

5. It is worth mentioning that Arendt is referring to the "the world-alienation of man of the modern age". Other concepts of alienation that have become widespread in modern social thought, such as "the disenchantment of the world" (Weber) and "alienation of man (Marx)", are perceived sharply critically by her on the grounds that they "often involve the romanticized notion of the past" (Arendt, 2006: 64). The question of whether the concept of world-alienation in Arendt's «The Human Condition» can be compared to the criticism of romanticizing ideas about the classical past is a matter for the readers to decide.

6. Cf.: Foucault, 2004.

modern world is that «the entire modern age has separated freedom from politics» (Arendt, 2020 [1960]: 37).

However, historical experience tells us that freedom and politics have not always been understood in this way. Arendt presents several arguments to demonstrate the close relationship between freedom and politics in the history of the Western world. Summing up, we can say that these arguments refer to «the oldest historical memories that have deposited themselves in our language, as well as the tradition of political thought and the experiences of the present» (Arendt, 2020 [1960]: 40), related to the experience of the right and left wing totalitarianism of the XX century. From a linguistic perspective, it is crucial for Arendt that «in all European languages we use a word for politics in which its origin, the Greek *polis*, can still be heard. Not only etymologically, and not only for scholars, this word is drenched with associations stemming from the community where politics in its specific sense was first discovered» (Arendt, 2020 [1960]: 37). Arendt also connects the idea that «tyranny is the worst of all forms of state» with the *polis* origin of both the term itself and politics as a free human activity. She believes that this idea is highly relevant for the self-understanding of statesmen and political theorists of the Western world in the second half of the 20th century. This unanimity is motivated by the fact that «among the classical forms of government tyranny is the only one that in principle cannot be reconciled with freedom» (Arendt, 2020 [1960]: 37 — 38). Consequently, this shared use of symbolic representations of the political realm by both the ruling elite and academic circles in the Western world prompts Arendt to suggest that the interests of security and reproduction of life, now encompassed in political thought by the concepts of "biopower" and "biopolitics," as introduced by Michel Foucault, cannot be the overarching priority in the interpretation of modernity's political objectives. After all, in this case we would have to admit that tyranny is not the worst way to govern the Earth, as it is possible to ensure safety and basic needs for those who are dominated. "If we really believed", Arendt points out, "as the theories of the modern age attempt to convince us, that in politics security and life interests are all that is at stake, we would have no reason to reject tyranny; for it can certainly deliver security, and it has often proved itself superior to all other forms of state in protecting mere life" (Arendt, 2020 [1960]: 38). In other words, «the original coincidence of freedom and politics, which was self-evident to classical antiquity» (Arendt, 2020 [1960]: 38) but then faded away, was partly preserved or revived in the symbolic understanding of Western modernity during the second half of the 20th century. All these arguments, taken together, offer an understanding of modern politics that not only "goes far beyond contemporary political theory and its conceptual framework", but also presupposes «a different consciousness of freedom and a different concept of politics to those we are accustomed to» (Arendt, 2020 [1960]: 41).

It is not without reason that Arendt places freedom at the heart of her understanding of man and politics. Human beings are unique creatures living in the world precisely because they have freedom. At the same time, Arendt does not understand freedom itself neither as freedom of will, in the sense of *liberum arbitrium,* or freedom to implement a rational project. Rather she sees it as an individual's capacity to transcend the "given" and

start something new. This capacity to start something new empowers human beings to liberate themselves from the constraints of imperatives of vital necessities and disrupt the inertial momentum of societal developments. This capacity to disrupt the automaticity of historical evolution and introduce something new into the world, from Arendt's perspective, is primarily connected with the realm of politics. It culminates in the historical experience of the great revolutions of modern age.

In the work «On Revolution» (Arendt, 1990 [1963]), whose key ideas are explained and clarified in the lecture "Freedom to Be Free", which eventually gave the name to the entire collection, Arendt settles accounts with two of the most influential intellectual traditions of the twentieth century — liberalism and Marxism. In her opinion, both these venerable traditions of political thought failed to fully understand the significance of the great modern revolutions — the American Revolution of 1776 and the French Revolution of 1789 — because they overlooked the central event: the establishment of a new public order based on political freedom, in which free and equal citizens could collectively deal with their common affairs. They failed to grasp the true *political* significance of modernity's great revolutions, because they viewed politics not as an independent and ultimate form of human activity, but rather as a means to achieve other, non-political objectives. This was the case for both liberals, who sought to pursue personal gain without bounds, and Marxists, who aimed to establish a classless society where citizens would be free and equal (Wellmer, 2006: 220).[7]

From Arendt's point of view, the main thing in the experience of modern revolutions is the coincidence of the idea of freedom with the beginning of something new: «the idea of freedom and the actual experience of making a new beginning in the historical continuum should coincide» (Arendt, 2020 [1966/1967]: 104)[8]. It is freedom, which is at the origin of the revolutionary movement that serves as an unmistakable criterion for distinguishing authentic revolutions from those that are not. For this reason, Arendt is clearly pleased to agree with Condorcet's words, who believed that the "word 'revolutionary' can be applied only to revolutions whose aim is freedom" (Arendt, 2020 [1966/1967]: 81). Arendt saw the main advantage of the American Revolution in that it not only declared independence from the British crown and founded a new state, but at the same time it was also able to establish a new political freedom space based on the 1787 Constitution, whose authors were inspired by the idea of separation of powers.

Speaking about the revolutions of modernity as an experience of a new beginning, Arendt first of all emphasizes the novelty of the modern interpretation of the concept of "revolution". Despite the fact that the word "revolution" can be used in a generic sense without taking into account either the word's origin or the temporal moment when the term was first applied to a particular political phenomenon, from a point of view of po-

7. Regarding the Marxist ideal of a classless society, Arendt in her notes from the 1950s rather sharply noted that Marx's vision of a society without classes, in which management of things would take the place of domination over people, was not apolitical but anti-political. Instead of not having any political government whatsoever, "it actually can only be rule by nobody, that is, bureaucracy, a form of government in which nobody takes responsibility" (Arendt, 2005: 77).

8. On the antinomies of this understanding of the idea of revolution in Arendt, see: Fine, 2001: 127 — 130.

litical philosophy a more differentiated approach is required that takes into account the term's historical context and its entry into the lexicon of modern political thinking. As Arendt emphasizes, "prior to the two great revolutions at the end of the eighteenth century and the specific sense it then acquired, the word "revolution" was hardly prominent in the vocabulary of political thought or practice" (Arendt, 2020 [1966/1967]: 82). In the seventeenth century the term is used in its original astronomical meaning, which signified the eternal, irresistible, ever-recurring motion of the heavenly bodies, meanwhile its political usage was metaphorical, describing a motion, a swinging back to a preordained order. Therefore, until and including the 18th century, this word-concept in the European cultural and civilizational area meant nothing more than "restoration", the content of which was recovery of former freedom, or rather, restoring of former liberties (Arendt, 2020 [1966/1967]: 83). This applies not only to the English Revolution of 1640-1660 and the "Glorious Revolution" of 1689, but also to the American Revolution of 1776 and the French Revolution of 1789. Thanks to the latter, which impacted the whole Europe, there was a significant change in the political meaning of the term, as a result of which it begins to denote the process of liberation of all people under the principles of liberty, equality and fraternity. Thus, Arendt concludes, «what actually happened at the end of the eighteenth century was that an attempt at restoration and recovery of old rights and privileges resulted in its exact opposite: a progressing development and the opening up of a future which defied all further attempts at acting or thinking in terms of a circular or revolving motion» (Arendt, 2020 [1966/1967]: 84).

Like the concept of revolution, the concept of freedom underwent similar, although more complex, substantial changes during the great revolutions of modern times. At the dawn of modernity, it was still identified with those rights and liberties we today associate with constitutional government and call civil rights. However, the problem was that those rights have nothing in common with the political right to participate in public affairs. Arendt attributes this to the fact that «liberties in the sense of civil rights are the results of liberation, but they are by no means the actual content of freedom, whose essence is admission to the public realm and participation in public affairs» (Arendt, 2020 [1966/1967]: 85). On this basis, Arendt draws a broad distinction between civil liberty and political freedom. While the former, based on civil rights, is inherently *negative* in nature, protecting human life from excessive interference by the state and society, the latter, in its essence, is a *positive* phenomenon associated with the right to participate in political decision-making about public affairs[9].

According to Arendt, it «was a passion for this new political freedom, though not yet equated with a republican form of government, which inspired and prepared those to enact a revolution without fully knowing what they were doing» (Arendt, 2020 [1966/1967]: 87). At the same time, Arendt sees the squaring of the circle of the great revolutions of modernity in the fact that, while they were associated with both liberation from the constraints of the "ancient regime" and the establishment of political freedom, the «lib-

9. For a discussion on the contradictions in Arendt's interpretation of the concepts of "freedom" and "liberty", see: Pitkin, 1988: 523 — 552, in particular 523 — 528.

eration is indeed a condition of freedom — though freedom is by no means a necessary result of liberation — it is difficult to see and say where the desire for liberation, to be free from oppression, ends, and the desire for freedom, to live a political life, begins» (Arendt, 2020 [1966/1967]: 86). If the desire to be free from oppression could have been fulfilled through the establishment of limited constitutional monarchy, then freedom as a political way of life requires the formation of a new or rediscovered form of government, namely, the constitution of a republic[10]. From Arendt's perspective, authentic revolutionary action aims to establish a constitution of freedom (*constitutio libertatis*), and the establishment of this constitution requires the formation of a republic that represents a constitutional government that guarantees not only basic civil rights but also the access of free and equal citizens to the political public sphere[11].

At the same time, Arendt reproaches the French Revolution, and with it, the symbolic universe that legitimized a specifically modern understanding of democracy's theory and practice, for putting a new absolute — the nation — in place of the old sovereign. This was a pernicious attempt to pour new wine into old bottles. In this case, we are dealing with the feature of the "modern social imaginaries"[12], which has long been observed by sociologists and social theorists, that «the old understanding of power and authority, even if their former representatives were most violently denounced, almost automatically led the new experience of power to be channeled into concepts which had just been vacated» (Arendt, 1990 [1963]: 155). And this, in turn, led to the deformation of new political concepts, forcing contemporaries to perceive and understand them in accordance with the patterns familiar from the old world. However, from Arendt' perspective, sovereignty and freedom in politics are not compatible. According to her, «where men, whether as individuals or in organized groups, wish to be sovereign, they must abolish freedom. But if they wish to be free, it is precisely sovereignty they must renounce» (Arendt, 2020: 57). This is another programmatic idea of Hannah Arendt — to create a democratic political theory that would be based not on the concept of sovereignty, whether it is the absolute sovereign of the "ancient regime" or a new sovereign represented by a single and indivisible modern nation.

Here we come to the most important — from the perspective of the relationship between philosophy and politics, thought and action — point in Arendt's political thought. The central place given by the author to the concepts of action, freedom, and revolution paints a normative picture of political life that is so at odds with the main trends in the development of the modern world and the prevailing approaches to these concepts that its implementation would, by the author's own admission, be a real "miracle" in practice. Consequently, numerous indications suggest that we are witnessing the emergence of a new secular redemptive theology, developed within the context of political philosophy,

10. Cf.: Arendt, 1990 [1963]: 33.

11. Some contemporary scholars believe that Arendt's position on *constitutio libertatis* is aporetic in nature, since "political freedom has an antinomical relation to the possibility of its own founding". This view is echoed in the work of Miguel Vatter, who interprets a political undertaking as a as a Machiavellian return to the origins (Vatter, 2000: 221).

12. The concept of "modern social imaginaries" is revealed in great detail today in the work of Charles Taylor: Taylor, 2004; Taylor, 2007.

where political action itself serves as the primary instrument of collective salvation, If not for free and equal citizens acting individually — death and life law will take its toll anyway — but at least for the *common world* as a public space where political freedom dwells. According to Arendt, «only the world and men in the plural can anticipate salvation through the miracle that is possible in all political affairs — at least as long as freedom, the human gift of interrupting ruin, remains intact. No miracle is required to save life as such since by nature it endures with the species, nor can a miracle ever save man in the singular, who must always die as an individual. These ruinous processes can be interrupted only for the world that is common to us all, which outlasts our life or at least can outlast it, and which is the specific concern of politics. From this it follows that, although the ability to begin may be a gift of man in his singularity, he can only realize it in relation to the world and in acting together with his fellow men» (Arendt, 2020 [1960]: 72 — 73).

It seems that within the framework of this new secular theology of redemption, the gift of freedom, the ability to begin and its revolutionary refractions as acts of forming something new and interrupting automatic inertial processes of social reproduction play a role similar to that played by "miracles" in religions of revelation. Where political action, as ability to begin and freely take initiative, disappears from society for one reason or another, then «the processes that freedom first brought forth also become automatic, and an automatic process produced by men is no less ruinous for the world than automatic natural processes are for the life of the individual» (Arendt 2020 [1960], 75). If it weren't for the miracle of freedom, as Hannah Arendt called it, then today, humans would be completely subordinate to these automatic processes that could only destroy their political existence in the world.

Therefore, as Arendt emphasizes, «this time, no less than the continued existence of men on earth may depend upon man's gift of performing "miracles," that is, bringing about the infinitely improbable and establishing it as a worldly reality» (Arendt, 2020 [1960], 76). In the context of this unorthodox view of the world and human action, Arendt and her followers can only hope that not only human freedom, as a way of being in the world, but above all, the gift of freedom itself, which was not created by humans but was given to them, will be preserved along with the opportunities they provide for revolutionary new beginnings: «without action, without the capacity to start something new and thus articulate the new beginning that comes into the world with the birth of each human being, the life of man, spent between birth and death, would inevitably be doomed beyond salvation» (Arendt, 2020 [1964]: 31). Thus, the promise of saving the common world of politics by word and deed through free human action becomes for Arendt not only an imperative for political life, but also a promise of salvation for both the humans and their freedom in the world.

References

Arendt H. (1978) *The Life of the Mind*, Vol. I. Thinking, N. Y.: Harcourt Brace Jovanovich.
Arendt H. (1979) On Hannah Arendt. *Hannah Arendt: The Recovery of the Public World* (ed. M. A. Hill), N. Y.: St. Martin's Press, pp. 301 — 340.

Arendt H. (1990 [1963]) *On Revolution*, L.: Penguin.
Arendt H. (1998 [1958]) *The Human Condition*, 2nd ed., Chicago: The University of Chicago Press.
Arendt H. (2005) *The Promise of Politics*, N. Y.: Schocken Books.
Arendt H. (2006) *Between Past and Future. Eight Exercises in Political Thought*, L.: Penguin.
Arendt H. (2020) *The Freedom to Be Free*, L.: Penguin.
Arendt H. (2020 [1964]) Labor, Work, Action. *The Freedom to Be Free*, L.: Penguin, pp. 1 — 32.
Arendt H. (2020 [1960]) Freedom and Action, A Lecture. *The Freedom to Be Free*, L.: Penguin, pp. 33 — 76.
Arendt H. (2020 [1966/1967]) The Freedom to Be Free. *The Freedom to Be Free*, L.: Penguin, pp. 77 — 110.
Fine R. (2001) *Political Investigations: Hegel, Marx, Arendt*, L.: Routledge.
Foucault M. (2004) *Naissance de la biopolitique. Cours au Collège de France (1978 — 1979)*, Paris: Gallimard; Seuil.
Lefort C. (1986) Hannah Arendt et la question du politique. *Essais sur le politique. XIXe — XXe siècles*, Paris: Éditions du Seuil, 1986, pp. 64 — 78.
Pitkin H. F. (1988) Are Freedom and Liberty Twins? *Political Theory*, vol. 16, no 4, pp. 523 — 552.
Taylor C. (2004) *Modern Social Imaginaries*, Durham; London: Duke University Press.
Taylor C. (2007) *The Secular Age*, Cambridge (Mass.): Harvard University Press.
Vatter M. (2000) *Between Form and Event: Machiavelli's Theory of Political Freedom,* N. Y.: Springer.
Wellmer A. (2006) Arendt on Revolution. *Cambridge Companion to Hannah Arendt* (ed. D. Villa), Cambridge: Cambridge University Press, pp. 220 — 244.

«Свобода быть свободным»: Ханна Арендт о «спасении» мира политического

Тимофей Дмитриев
Кандидат философских наук, доцент школы философии и культурологии
факультета гуманитарных наук
Национальный исследовательский университет «Высшая школа экономики»
Адрес: улица Старая Басманная, 21/4с1, Москва, 105066 Российская Федерация
E-mail: tdmitriev@hse.ru

Публикации последних лет и десятилетий показывают, что помимо написания монографических исследований и издания сборников статей, значительная часть которых была первоначально опубликована в различных гуманитарных изданиях, Ханна Арендт (1906 — 1975) на протяжении всей своей академической карьеры в США активно использовала также жанр отдельных лекционных выступлений для разъяснения и популяризации идей своей политической философии и своего подхода к актуальным

проблемам современности. Их публикация позволяет во многом по-новому взглянуть на важные аспекты ее представлений о мире политического и о роли политической свободы, преломляющейся в словах и поступках людей в этом общем для них мире. В центре внимания данной статьи находятся лекции Арендт 1960-х годов, посвященные прояснению замысла двух ее ключевых политических работ второй половины 1950-х — начала 1960-х — «The Human Condition» (1958) и «On Revolution» (1963), — а также интерпретации центральных для дискурса ее политической философии понятий действия (Action), свободы (Freedom), политики (Politics) и революции (Revolution). Особое внимание в статье уделяется предложенному Арендт в этих лекциях толкованию свободы человека как дара (gift) и чуда (miracle), негативная работа которой в мире человеческих слов и дел заключается в прерывании автоматизма инерционных процессов социального воспроизводства, а позитивная — в «спасении» (salvation) общего мира политического для ныне живущих и будущих поколений.

Ключевые слова: Ханна Арендт, политическая философия, действие, свобода, мир политического, революция, спасение.

doi: 10.17323/1728-192x-2024-4-259-266

"…Such is the Reality of International Politics"

Book review: Mearsheimer J. J., Rosato S. (2023) How States Think: The Rationality of Foreign Policy, New Haven and London: Yale University Press. — 304 pp. ISBN 978-0-30026-930-7.

Maxim Fetisov
Candidate of Philosophical Sciences
Independent Researcher
Address: Lermontova str., 241, Stavropol, 355041 Russian Federation
E-mail: msfetisov@gmail.com

> Some will remind us of the saying "while the Romans deliberate, Saguntum is lost." On the other hand, when the few decide everything, simply on the basis of their own affects, freedom and the common good are lost. For human wits are too sluggish to penetrate everything right away. But by asking advice, listening, and arguing, they're sharpened. When people try all means, in the end they find ways to the things they want which everyone approves, and no one had ever thought of before.
> Spinoza, Political Treatise, ix, 14

The realist perspective on global politics, which once existed in the shadows of ideological euphoria, is now experiencing a remarkable resurgence, albeit primarily in practical terms. As is often the case in such situations, its conceptual and doctrinal design is somewhat delayed. The old realist authors were seen as relics of the Cold War, preserved in its remote corners, somewhere in time, while the departments of international relations, dominated by ideological triumphalism, failed to bring up the new generations of realists for academia. Thus, by the second decade of this century, a meeting with a foreign policy realist equals an encounter with a prehistoric creature.

John J. Mearsheimer, the R. Wendell Harrison Distinguished Service Professor at the University of Chicago and a renowned scholar of international relations, has always been an exception to this rule. As a foreign policy theorist, he never stepped back, retreating into liberal fairy tales about "democratic peace", "rules-based order", or "decline of violence". He always firmly adhered to the vision of the political world established by Thucydides, Machiavelli, Hobbes and Hans Morgenthau. This is a world riddled with chaos, anarchy, and ignorance; here, "the large ones eat the smaller" and your survival depends on the actual ratio of power and strength. This is the case that Mearsheimer makes in his books and in numerous public appearances. The main players in this scenario are still the states, and as we know from the classics of political philosophy, the virtue of a state is security. This is not a particularly promising outlook, so is it even possible to have any stable political knowledge within this grim perspective? The question is discussed in the book "How States Think: The Rationality of Foreign Policy" that was written by Mearsheimer together with his former student Sebastian Rosato, who is a political science professor at the University of Notre Dame.

The general answer is yes, because if this were not the case, then no meaningful statements about international politics would be possible at all: "If nonrationality is the norm, state behavior can be neither understood nor predicted, and studying international politics is a futile endeavor. For practitioners, rationality enables states to devise effective foreign policies. Only if those other states are rational actors can one anticipate how friends and enemies are likely to behave in a given situation and thus formulate policies that will advance one's interests" (p. xiv).

Moreover, according to the authors, most foreign policy actors are primarily concerned with trying to behave in the most rational manner. This line of argument runs contrary to the familiar trope of Western discourse that habitually accuses opponents of "irrationality" or even insanity. This perspective does afford us the opportunity to perceive certain events in a fresh light, particularly in the context of crisis and warfare.

The authors begin their argument by covering current events: "It is widely believed in the West that Russian president Vladimir Putin's decision to invade Ukraine was not a rational act." Putin's critics also insist that "the only morally acceptable reason for going to war is self-defense, but the invasion of Ukraine was a war of conquest" (p. ix). The main reason that prompted Putin to launch the Ukrainian campaign, they say, is that "Putin was bent on conquering Ukraine and other countries in Eastern Europe to create a greater Russian empire, something that would satisfy a nostalgic yearning among Russians but that makes no strategic sense in the modern world" (p. xi).

This is the point that Mearsheimer and Rosato strongly disagree with. They say that these "claims rest on common understandings of rationality that are intuitively plausible but ultimately flawed. Contrary to what many people think, we cannot equate rationality with success and nonrationality with failure. Rationality is not about outcomes. Rational actors often fail to achieve their goals, not because of foolish thinking but because of factors they can neither anticipate nor control. There is also a powerful tendency to equate rationality with morality since both qualities are thought to be features of enlightened thinking. But that too is a mistake. Rational policies can violate widely accepted standards of conduct and may even be murderously unjust" (p. x).

But actual situation, in the opinion of the authors, was precisely the opposite. Russia's decision to invade Ukraine was rational because "Putin and his advisers thought in terms of straightforward balance-of-power theory, viewing the West's efforts to make Ukraine a bulwark on Russia's border as an existential threat that could not be allowed to stand… In short, this was a war of self-defense aimed at preventing an adverse shift in the balance of power." (p. xi-xii). Moreover, as the evidence indicates contrary to all widely shared rhetoric about "Putin's reckless bid for empire", Russia's decision to openly intervene in the civil war in Ukraine was not only a product of some sort of strategic rationality, but also the result of a lengthy deliberative process: "Putin's subordinates shared his views about the nature of the threat confronting Russia, and he consulted with them before deciding on war" (p. xiii).

If this is the case, then the decision to start a war by invading Ukraine was not only rational, but there was nothing unique or anomalous about it. Moreover, history pro-

vides us with a lot of examples of such actions by great powers, when their seemingly irrational behavior actually turned out not only to be rational but was taken after a due deliberation process (p. xiii). It was the discrimination and criminalization of war that occurred after 1945, its ban from the list of possible ways to resolve political disputes, that made it possible to perceive the state's decision on war as deeply irrational. It is this widely shared delusion that Mearsheimer and Rosato seek to unravel: "Against the increasingly common view among students of international politics that states are often nonrational, we argue in this book that most states are rational most of the time" (p. xiv). The initial hypothesis put forward by the authors proceeds from the fact that there is no "good definition" of rationality in the international relations scholarship. They suppose it something to be about "making sense of the world world — that is, figuring out how it works and why — in order to decide how to achieve certain goals. It has both an individual and a collective dimension. Rational policymakers are theory-driven. They rely on credible theories — logical explanations based on realistic assumptions and supported by substantial evidence — about the workings of the international system, and they employ these to understand their situation and determine how best to navigate it. Rational states aggregate the views of key policymakers through a deliberative process, one marked by robust and uninhibited debate. In sum, rational decisions in international politics rest on credible theories about how the world works and emerge from a deliberative decision-making process" (p. x-xi).

However, the vast majority of the research literature on foreign policy proceeds from completely different premises, which the authors find completely unsatisfactory and deeply flawed. There is a general framework for international relations scholarship, this is the so-called "rational actor assumption" (p. 1). It implies that key decision-makers act, as a rule, in a rational way. This means that they make sense of the world by making up credible theories about its workings and composition. This assumption "has both individual and state-level dimensions". When describing the individual dimension, the authors resort to the help of Max Weber (albeit through the mediation of Steven Kalberg): "mental processes that consciously strive to master reality are common to all the types of rationality. . . All of these processes systematically confront… social reality's endless stream of concrete occurrences, unconnected events, and punctuated happenings. In mastering reality, their common aim is to banish particularized perceptions by ordering them into comprehensible and 'meaningful' regularities" (p. 20-21). This is quite a traditional definition, however, when shifting to the collective dimension, it takes on a certain intellectualist and deliberative twist: "A state is rational if the views of its key decision makers are aggregated through a deliberative process and the final policy is based on a credible theory. Conversely, a state is nonrational if it does not base its strategy on a credible theory, does not deliberate, or both. A careful review of the historical record shows that judged by these criteria, states are regularly rational in their foreign policy" (p. 2). The authors assume this hypothesis to be a "radical intervention in the debate" since they are going to "offer a meaningful definition of rationality in international politics where none existed" (p. 4).

It is important to make a clear distinction between two different types of rationality: the rationality of means and the rationality of ends: "there is a difference between what we call 'strategic rationality' and 'goal rationality'". It is the distinction that is most wanting in contemporary foreign policy scholarship since "the debate on rationality in the international relations literature focuses almost exclusively on whether a state's strategies are rational and pays little attention to evaluating the rationality of its goals (p. 4).

There is an old philosophical habit of attributing the strategic rationality of means to the sphere of instrumental action, which is controlled by a calculated cost-benefit ratio aimed at solving a specific problem. The rationality of goals, on the other hand, is linked to the realm of morality and is governed by ethical imperatives. This leads to obvious confusion and misunderstanding when decision-makers, political scientists, journalists, and even historians on numerous occasions describe the policies of their opponents or subjects as "nonrational". Consequently, when they confuse morality with political affairs, they become susceptible to cognitive biases. This old dichotomy, however, does not accurately express the situation in foreign policy analysis. According to the realist logic, the highest priority of the state dealing with the international system is survival. This means that all strategic policies and subsequent actions should be entirely subordinated to this ultimate goal: "survival is particularly important. States aim to preserve the integrity of their physical base and maintain their ability to determine their own political fate" (p. 213).

The paramount importance of survival necessitates that foreign policy must be proactive in the face of high uncertainty, which is a defining and essential feature of interstate relations, with war being the most critical factor: "Policymakers confront serious information deficits regarding most of the elements that matter for designing grand strategies or navigating crises. The farther they peer into the future, the larger these deficits become. Among other things, policymakers may not have good data about their own people's resolve or how their weaponry and combat forces will perform in a war. Additional uncertainties apply when it comes to assessing other states, friends as well as enemies" (p. 25).

It is important to differentiate between uncertainty and risk, which is simply choosing from a range of pre-determined options: "In a risk world, decision makers do not know the consequences of pursuing any given strategy, but they can acquire the information needed to calculate the odds of various outcomes… In an uncertain world, actors cannot acquire the information needed to evaluate the likely consequences of pursuing different strategies." (p. 23-24).

Rational decision-making in a situation of uncertainty requires not only collecting and processing vast amounts of raw data, but also doing so in a fundamentally information-deficient environment, such as international politics: "It is difficult to measure the military assets, objectives, intentions, and strategies of other states, especially since states often conceal or misrepresent their capabilities and thinking. Taken together, these information deficits mean that decision makers are bound to have limited knowledge about

how their states' interactions with other states are likely to play out and to what outcome" (p. 25)

This data-driven approach does not constitute a rational policy in itself. The transition to rational action requires a reliable theory supported by an active process of deliberation: "A state's policy is rational if it is based on a credible theory or some combination of credible theories and is the product of a deliberative process. Policies that do not rest on credible theories or are not the products of a deliberative process are not rational" (p. 65).

The bulk of ideas about rationality in foreign policy can be divided into two main paradigms: the rational choice theory and behavioral economics. Both of these patterns, for all their differences, proceed from the axiom of expected utility maximization: "There are two bodies of scholarship that explore the rational actor assumption in international politics. Rational choice scholars and political psychologists both think about rationality in terms of expected utility maximization, which is basically a data-driven enterprise. But they emphasize different issues: the former focus on rationality while the latter focus on nonrationality" (p. 70). The actor in this scenario is homo economicus, aiming not for survival, but for maximizing profit. This kind of behavior is considered to be rational. At the same time, rational choice theories focus on outcomes and do not describe the actual process of making decisions. In other words, they do not describe the workings of rationality. Mearsheimer and Rosato even introduce the Kantian term "as if" in their description. They say that the aforementioned researchers "pay scant attention to how rational policymakers make sense of the world or how rational states aggregate the views of those individuals. They do examine how individuals make choices, as one would expect from scholars who describe themselves as rational choice theorists. Yet they do not discuss the mental processes by which rational policymakers make decisions. Instead, they assume that those individuals act "as if" they were expected utility maximizers" (p. 70)

Behavioral economics, despite the fact that it describes the work of rationality from a different perspective, also relies on the figure of homo economicus. The essential difference is that it tries to describe the actual workings of rationality in foreign policy, but it does so by identifying various cognitive tools that are explicitly or implicitly used in decision-making: heuristics, shortcuts, historical analogies, and so on (p. 91). Thus, it identifies various errors, cognitive biases, i.e., rationality limitations, in order to come to the conclusion that there is no rationality in world politics and most of the decisions made are essentially nonrational: "Political psychologists define nonrationality as deviation from expected utility maximization, which they call bias. Focusing almost exclusively on how individuals make choices, they argue that policymakers routinely rely on mental shortcuts — primarily analogies and heuristics — that lead to biases" (p. 71).

From the perspective of Mearsheimer and Rosato, political psychology and behavioral economics seem to be in a clear performative contradiction, since if their arguments were true, no international policy could be implemented at all. Nevertheless, the tools of political psychology (hard-wired into the human brain or based on observations of historical events) and behavioral economics are actively used by political scientists, experts, and journalists. This leads to another source of cognitive errors and unacceptable confu-

sion between politics and morality, as is regularly demonstrated by impressive failures in various foreign policy areas.

It is also important to note that neither rational choice theory nor political psychology, with behavioral economics, provide a consistent perspective for understanding the process of aggregating individual decisions into collective strategies. In their view, the actions of individuals and the actions of the states look equivalent, which is of course not the case in reality (pp. 73-74).

This, of course, does not mean a complete absence of nonrationality in international affairs. However, it needs to be assessed within a completely different conceptual framework. To support their argument, Mearsheimer and Rosato provide a series of historical examples to illustrate the process of developing grand strategy and managing crises. The authors demonstrate that these events, traditionally viewed by political scientists and historians as nonrational and insane decisions, were actually well-reasoned attempts to implement rational policies in situations of high uncertainty. Such are, for example, Germany's decision to enter the Great War in Europe, The United States' decision to pursue liberal hegemony after the Cold War, Japan's decision to attack Pearl Harbor, the settling of the Cuban Missile Crisis or Soviet decision to invade Czechoslovakia (pp. 103-179). These cases are clearly in line with the criteria for rational state behavior in international politics outlined by Mearsheimer and Rosato: the availability of a reliable theory and a robust deliberation prior to decision-making. The mere existence of rationality does not guarantee success in foreign policy. There are numerous other factors, such as insufficient data, misinterpretation of intelligence, and more, that can influence the outcome: "policies derived from credible theories sometimes fail because circumstances change in important and unexpected ways — what Niccolo Machiavelli calls fortune and both Thucydides and Clausewitz call chance" (p. 68). However, the authors propose assessing not the outcomes of decisions but the processes that lead to them. Conversely, there are nonrational ways of acting and making decisions. The authors explore several such examples: Germany's decision on a risk strategy before the World War I, Britain's no-liability strategy prior to World War II, the United States' bid to invade Cuba in 1961, and the United States decision to invade Iraq in 2003 (pp. 180-209). Although nonrational policies can sometimes lead to success, it depends on factors that are unrelated to the quality of the decision-making process. The criteria of nonrational state behavior are the ban or exclusion of information necessary for forming reliable theories, as well as absence or distortion of proper deliberation.

A full understanding of state strategic rationality is impossible outside the broader framework, namely the goals for which states exist: "if rationality means making sense of the world for purposes of navigating it in pursuit of particular goals, then an understanding of the concept must involve how states think about their goals as well as how they pursue them" (p. 211). Here, the authors repeat their "bottom-up" logic of political action, again referring to Max Weber (and again through Kalberg): "Something is not of itself 'irrational,' but rather becomes so when examined from a specific 'rational' *standpoint*" (p. 212). A "rational standpoint" that determines all other goals and actions is, as we men-

tioned earlier, survival. It is an idea that comes well from Thomas Hobbes, who maintains that "reason" tells us that "a man is forbidden to do that which is destructive of his life, or takes away the means of preserving the same" (p. 213).

The same applies to the state: "There is only one inviolable rule. Survival is primary, and all other objectives must be subordinated to it. It is a matter of incontrovertible logic and evidence that a state cannot achieve any other goal if it does not first survive as a state" (p. 213). The state, according to the authors, drawing on the work of the sociologist Charles Tilly, represents a historical form of human survival: "Human beings — who prize survival above all other goals — are social animals. They are born into and operate in tight-knit social groups, which also rank survival as their number one goal. To function effectively and protect their constituents, these groups construct political institutions". These institutions, i.e., "states", have existed throughout f human history, but retained their primary goal (p. 214).

Of course, states may put this primary goal at risk, but this risk does not arise from the fact that survival is not prioritized or considered subordinate compared to other goals. Instead, it is usually associated with overstretched power, which occurs when there is an incorrect identification of threats or excessive involvement in security competition, as was the case with Germany's engagement in the general European war in 1914, or with Japan's decision to launch Word War II in the Pacific (p. 219). Foreign policy, therefore, is a dangerous undertaking and one should not confuse politics with morality: "Rational decision makers simply try to figure out the most effective strategy for dealing with other states, and as should be apparent by now, threatening or initiating violence sometimes makes sense. This message is hardly uplifting, but such is the reality of international politics" (p. 225).

The attempt made by Mearsheimer and Rosato in their book to protect foreign policy analysis from moral reasoning, to prevent the invasion of economics into political theory, and to restore its independence is interesting in itself and deserves attention. Nevertheless, it does not dismiss certain problems. The prioritization of survival, for example, which is unconditional in relation to other goals, undoubtedly works in a system of foreign policy based on interaction between nation states. But what will be the hierarchy of goals in the case of some "post-national constellation" or "after empire"?[1] Assuming, of course, that such conceptual constructions pass the reality check. There are certain doubts about this.[2]

The description of a political animal as homo theoreticus also raises the question of the gap between theory and practice. This intellectualism is not new — it is a well-known problem that has been around for a long time. The hypothesis suggests that if something makes sense at the level of theory, then its practical implementation is not bound to come across any obstacles. However, this is often not the case. Just as any war plan does not survive its first meeting with front lines, so the implementation of any seemingly

1. Habermas J. (2001) *The Postnational Constellation: Political Essays*, Mass.: MIT Press.

2. *Certain objections to Mearsheimer's argument see in Todd E.* (2024) La Défaite de l'Occident, Paris: Gallimard.

perfect political idea often encounters resistance that either destroys it or alters it beyond recognition.

It is this gap that we mean when we talk about, for example, of the role of affect in politics. What might the deliberation theory look like in that case? The very concept of deliberation bears distinct Habermasian turn that implies non-manipulated, unbiased, "domination-free communication". However, is this the case with actual deliberation when making a political decision? Abundant historical evidence suggests otherwise. Could this absence of communicative reciprocity be accounted for by theoretical frameworks? How to address the issue of popular sentiment? It is evident that no economic framework for political action can fully explain it. But what about the intellectualist models, the advocates of which the authors claim to be? Can they do? The authors explicitly deny any theoretical status for affect, although it is clear that Mearsheimer fully understands the significance of collective feelings, and in his previous work, for example, he gave sufficient attention to nationalism as the most significant driving force in world politics.[3] The authors mention the importance of the affective dimension in politics and even refer to Antonio Damasio, a renowned neuroscientist (p. 60). Damasio is famous for his "Spinozist turn" in neuroscience that showed an inextricable link between rationality and affect.[4] The development of this topic is wanting in the book, although it is clear that the authors had a different initial purpose. Hopefully, this void will be filled in the future.

«…Такова действительность международной политики»

Рецензия на книгу: Mearsheimer J. J., Rosato S. (2023). *How States Think: The Rationality of Foreign Policy*. New Haven and London: Yale University Press. — 304 pp. ISBN 978-0-30026-930-7.

Максим Фетисов
Кандидат философских наук, независимый исследователь
Адрес: ул. Лермонтова, 241, г. Ставрополь, 355041 Российская Федерация
E-mail: msfetisov@gmail.com

3. Mearsheimer J. (2018) *The Great Delusion: Liberal Dreams and International Realities*, New Haven and London: Yale University Press.

4. Damasio A. (2003) *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*, London: William Heinemann.

doi: 10.17323/1728-192x-2024-4-267-273

In Search of the Lost but Much-Needed Peace: The Origins and Dynamics of the Second Cold War

Book review: Richard Sakwa (2023) *The Lost Peace: How the West Failed to Prevent a Second Cold War*, Yale University Press: New Haven & London.

Stanislav L. Tkachenko
Doctor of Science (Econ.), Professor at the Department of European Studies, Director of the MA program "Diplomacy of Russian Federation and Foreign States", School of International Relations,
Saint-Petersburg State University
Address: Entrance 8, 1-3 Smolny street, Saint-Petersburg, 191060, Russian Federation
E-mail: s.tkachenko@spbu.ru

The author of the book under review, British political scientist Richard Sakwa, is well-known in academic circles and the expert community worldwide. At the beginning of his academic career, he taught Soviet and European politics at the University of California in Santa Cruz (USA) and the University of Essex (UK). For more than 30 years, he was a professor at the University of Kent, where he also served as head of the Department of Politics and International Relations. Professor Sakwa is the author of two dozen monographs and more than a hundred articles on the problems of international relations and European security, Soviet and Russian history, a participant in the Valdai International Discussion Club and an expert at the Royal Institute of International Affairs (Chatham House) in London.

Published by the Yale University Press in 2023, 'The Lost Peace: How the West Failed to Prevent the Second Cold War' is the result of Richard Sakwa's long-term research into the process of formation, adaptation, and transformation of the contemporary architecture of international security system. Over the past three decades, the European model of international relations has seen both periods of unbridled optimism about its long-term sustainability and appeal to other regions of the world, as well as bitter disappointments regarding its ability to maintain peace and stability in Eurasia. Divided in 1945 by the Iron Curtain, Europe has come a long way from finding hope for peace in the late 1980s to the loss of this illusion and the growing threat of global military conflict in the early 2020s. The author focuses primarily on the historical period after the collapse of the USSR, while also referring to earlier events in Soviet, Western European, and American history. The monograph summarizes the findings of the author's previous scholarship and makes predictions about the future development of the international security system in the near future.

The monograph under review is divided into three main parts.

In the first part, *From Cold War to Hot War*, Richard Sakwa assesses the debate around the "peace dividends" that emerged followng the end of the bloc confrontation. He explains in detail, based on a study of documentary sources and memoirs, how internal instability in the USSR and the M. Gorbachev's weakness as a negotiator led Western leaders to conclusde that it was possible to obtain unilateral benefits from the end of the Cold War and exclude the USSR from its beneficiaries. The point of no return for the European

system of international relations was the decision of the Clinton Administration in 1994 to expand NATO into the states of Eastern Europe, including the former Soviet Baltic Republics. The author has compiled a list of about 30 statements made by Western leaders made at the turn of 1989-1990, which asserted that NATO would not expand eastward in response to the dissolution of the Warsaw Pact. The reversal of this position, which led to the current severe crisis in Europe and 'The Lost Peace' momentum, was due to Washington's conviction that a neutral unified Germany would inevitably lead to NATO's dissolution, since the alliance would lose its original mission and be unable to formulate a new one. However, at the turn of the 1980s and 1990s, instead of making a decision to dissolve NATO, which was fair from Moscow's point of view, Washington announced that the North Atlantic Alliance had become the most successful military alliance in world history; it had defeated the USSR during the Cold War and was prepared to solve new global problems. A popular statement of those years, made by Senator Richard Lugar on August 2, 1993, was: "NATO should go out of area or out of business". This 'out of area' movement has led Russia and NATO to the current standoff in Ukraine, where Russia argues for the preservation of the previous international law system based on the UN Charter (the author calls it 'the Charter International System'), while the United States tries to prove the viability of an alternative model called the 'Liberal Universalist System.'

The second part of the book, *Great-Power Conflict*, focuses on how the major global policy players (the United States, China, Russia, and the European Union) formulate and implement their views of the optimal world order. Particular attention is paid to the genesis of the US global strategy, which rejects the supremacy of traditional 'international law' in favor of 'rules-based' order. According to the author, this step signals Washington's rejection of the previous treaty-based, impartial system of international law and its commitment to unlimited efforts to promote its own interests on a global scale. The other three 'major world league' players (China, Russia, the EU) are forced to respond to US actions in accordance with their national traditions and current capabilities. Thus, Russia, which R. Sakwa describes as a neo-revisionist power (p. 174), has expressed its full-fledged rejection of Washington's destruction of the international legal system. China is focused on building its economic power and technological decoupling with the aim of potentially engaging in a conflict with Washington in the future. Finally, the European Union, due to its internal weakness and increasing fragmentation, is forced to follow in the wake of American foreign and defense policies towards consolidated Atlantism.

Finally, the third section, *War and International Politics*, is devoted to the dismantling of the strategic nuclear arms control regime and the current escalation of the conflict between Russia and the West. The stages of this process were the "Crimean Spring" of 2014 and the Special Military Operation of 2022. According to Professor Sakwa, the result of the largest military conflict in Europe since 1945, currently underway in Ukraine, has been an unprecedented crisis in the international security system and the threat of a global nuclear conflict. At the same time, processes that are not planned by Western countries and are undesirable for them are emerging. These include the rise of what the author refers to as the "Political East" with the strengthening of the role of China, India

and other countries of the global South in international affairs. Another example is the emergence of a new type of "global majority", an association of countries that do not see a place for themselves in the American-centered global economy and security model. In conclusion, the author expresses the hope that thanks to the efforts of Russia and other countries that share its vision of the world, it is the Charter System that will ensure stability in a new multipolar world rather than an "international liberal rules-based order".

The theoretical framework of the study can be described as eclectic. Such a complex topic as the collapse of the previous world order and the prospects for the formation of a new international political and economic system requires the study of a wide range of state and non-state actors, analysis of hard and soft security issues, challenges to economic development, functioning of multilateral institutions, and assessment of key features of political economies and social movements in Russia and other states. Consequently, in the text of the book, we can observe the implementation of the realist perspective in international relations (neo-realism, structural realism, the theory of hegemonic stability, offensive realism), the liberal perspective (the liberal internationalist school, multilateral governance/multilateralism), and the constructivist perspective (the Copenhagen School).

R. Sakwa summarizes his views on the theoretical foundations needed to built a new system of international relations with the concept of sovereign internationalism. He defines it as follows: "Sovereign internationalism represents an alternative to both neorealism, with its emphasis on the balance of power, spheres of interest, balancing and the like, and to fully fledged liberal internationalism, which includes a whole range of other attributes, including free trade and liberal democracy" (p. 18)

The author also brings back the characterization of international relations as a sphere of 'power politics', popular with neo-realists, especially John Mearsheimer (p. 4, 22), into the academic discussion of the reasons for the collapse of the previous world order model. The term 'power politics' emerged at the center of discussions about the nature of international relations on the eve of World War II thanks to the book by British scholar and professor at University College in London,[1] Georg Schwarzenberger . This term became widespread during the Cold War as a description of the goals of states and the methods they use in their activities on the international stage. The scholars who apply the concept of 'power politics' to the studiy of international relations assume that it can be effectively used to separate the international arena from the domestic political space. In this space power (coercion) has subordinate importance compared to existing legal norms, including national constitutions, laws and by-laws, as well as decisions of authorized bodies of public administration.

The author formulated the following three research questions for his monograph:

1) Did the reduction of ideological differences between the superpowers during the final stages of the Cold War help create conditions for building lasting peace?

2) Could the hypothetical de-ideologization of interstate relations under new conditions lead to the restoration of independence for multilateral cooperation structures, such as the UN and CSCE and overcome the division into opposing blocs?

1. *Schwarzenberger G.* Power Politics: An Introduction to the Study of International Relations and Post-war Planning. J. Cape, 1941.

3) Were conditions created in which relations between great powers aimed at achieving the goals of international security and development? Were they doomed to conflict and the reproduction of hierarchical power structures?

According to R. Sakwa, today we know that the answer should be negative in all three cases. His study aims to explain the reasons for this disappointing development of events and also highlights the factors that prevented the consolidation of the ideals of a "positive peace", which gave way to another round of conflict between the great powers, known as the "Second Cold War".

R. Sakwa's thesis that the main feature of the era that began in 1945 is 'the Charter international system', invented and established by Allied Powers is scientifically novel and makes an important contribution to the discussion on the genesis of the international system after the end of the World War II. In circumstances where the USSR could not agree with its partners in the anti-Hitler coalition that the unconditional surrender of Nazi Germany should be followed by a "peace treaty" along the lines of the Versailles Peace of 1919, it was the UN Charter that established the rules of conduct in the post-war international arena. It ensured an unprecedentedly long period of peace on the European continent, established favorable conditions for the disintegration of the European colonial system and provided a smooth transition of the permanent seat in the UN Security Council from the USSR to the Russian Federation.

According to R. Sakwa, the USSR under the protection of the Charter international system received a legal basis for sovereign internationalism and ultimately for multipolarity as a diversity of countries and social systems. Moscow acquired the opportunity to protect loyal members of international community through multilateral diplomacy (p. 318). On the other hand, the United States was able to put into practice the idea that the freedom of choice of the states, oriented towards Washington, can be implemented even if it threatens the security of other countries. Thus, at the peak of the Cold War, conditions were set for American global supremacy and the liberal transformation of the world that followed in the 1990s. Richard Sakwa correctly notes that Mikhail Gorbachev's calls to strengthen the Charter international system could not be heard in Washington, as they contradicted the U. S. model of liberal internationalism. At the level of political and diplomatic discourse, a confrontation emerged between Gorbachev's vision of a 'Common European Home' and the United States' narrative of 'common ideals', on which an alliance of shared values was supposed to be built.

The decision of the Clinton administration to expand NATO to Eastern Europewas driven by a set of geopolitical considerations. It became Washington's first major step towards building a system of institutions for a unipolar world, based on the rules of hegemonic world order (p. 6) and the convictions of hegemonic stability theory, which gives the leading state the mission to buildi and maintain an open and liberal world economy. (p. 116) Richard Sakwa is skeptical about the sustainability of the "dual hegemony" system, military and economic at the same time. This hegemony-leadership requires from the United States, as the world's most powerful country, not only immense resources to

create and maintain it, but also a willingness to sacrifice its own interests in favor of the abstract principles of the "common good" and the benefits of long-term stability.

At the final stage of the Cold War, the struggle between the USA and the USSR for the right to set world order agendas intensified. As the author notes, Washington did not support M. Gorbachev's expectations for abandoning bloc thinking and building interstate relations based on the principles of trust and pluralism. These ideas were not alien to the states of Western Europe during the period under review. But, having found itself in the process of transforming its initially predominantly economic association into a political union with supranational governance institutions in the late 1980s, the renewed and expanded European Union did not support Gorbachev's pan-European agenda. The author's conclusion that the USA did not plan to work for the "common good" at the final stage of the Cold War, but acted in their own selfish interests, seems justified to us.

Of the many aspects of the crisis in European security system, R.Sakwa identifies three, in our view, as the most important.

The first is the confrontation between the liberal international order and sovereign internationalism during the Cold War. The former was supported by the United States and its allies in Europe and beyond, while the latter was supported by the USSR and China. We agree with the author's thesis that the ideological origins of liberal internationalism go back to the Enlightenment with its ideals of progress, rationalism, free trade, and interstate cooperation. According to Richard Sakwa, the two basic pillars of the liberal international order are: 1) An open trade and financial system embodied in the GATT/WTO alliances and the Bretton Woods institutions; 2) An extensive military infrastructure, which began to be constructed in April 1949 with the signing of the Washington Treaty and establishment of the NATO military bloc under US hegemony. As the author correctly notes, during the first Cold War, the term 'liberal' should have been interpreted as "anti-communist," and not at all "liberal democratic," as is commonly thought today. The liberal international order should therefore rightly be viewed as a 'hegemonic international order' led from Washington (p. 6). The opposing sovereign internationalism was weakened during the period under review due to the China-Soviet standoff, the socialist countries' lag in the scientific and technological revolution, and the rapid development of post-industrial society. The USSR and its allies were characterized by a conviction that the protection of human rights belonged to state structures, rather than to institutions of society operating within or outside national borders. In the final stages of the Cold War, sovereign internationalism was rapidly losing its adherents and was no longer able to offer anything attractive in the face of the pressure of the liberal international order.

The second aspect is the failure to anticipate the establishment of what Sakwa calls a 'positive peace order'. He defines it as follows: "A positive peace order in our case is one in which the actors cooperate within the framework of the broader international system guided by the principles of sovereign internationalism and international law" (p. 10). According to the author, this order goes back to the idea presented by US President J. F. Kennedy in his speech at the American University in Washington in June 1963: 'peace is a process — a way of solving problems'. The tragedy of the post-Cold War period in

world politics is that this process never began. The practice of negative peace, focused on conflict *management* rather than conflict *resolution*, continued to dominate interstate relations. But, as the current proxy war between Russia and the political West in Ukraine shows, even conflict management according to the recipes of the Cold War ultimately proved to be ineffective.

The third aspect is the triumph of Atlanticism over pan-continental Europeanism in the contemporary period of history. Endless debates about whether the leaders of the United States and Western European countries promised Gorbachev that NATO would not advance eastward even an inch are incidental evidence of a more significant phenomenon: the non-negotiated settlement of the Cold War in 1989-1991. The end of World War II was formalized, at least, in the form of the UN Charter, which has been the basis for the entire system of international law and global political relations has been built for over 70 years. The end of the Cold War, however, was only formalized in the form of several non-binding declarations. A perfect example is the Charter of Paris for a New Europe, approved on November 21, 1990 by the leaders of 34 states participating in the Conference on Security and Cooperation in Europe. Later, the Charter marked the beginning of the institutionalization of the CSCE, leading to its transformation into the Organization for Security and Cooperation in Europe (OSCE) and the creation of several permanent bodies, including the OSCE Parliamentary Assembly and the OSCE Office for Democratic Institutions and Human Rights (ODIHR). Still, the influence of the OSCE on processes in real European politics has turned out to be minimal, allowing Atlanticism to survive and even strengthen in this century.

The "Lost Peace", studied by Professor Sakwa was the world of unilateral concessions by Gorbachev to Ronald Reagan and George Bush Sr. It was a world of unjustified expectation in Moscow that there would be a change in the nature of international relations and a move towards universal harmony. Can this temporary and obviously artificial state of affairs between the two superpowers and their leaders be described as "peace", the crisis and subsequent collapse of which can be characterized as a "loss"? This question, important for understanding the historical process in the 1980s, remains unanswered in the book. In general, one should agree with the author's opinion that the decision of the Allied Powers in World War II to limit themselves to adopting the UN Charter rather than a full-fledged peace treaty was a forced and erroneous one. Their hopes that they could thus "kill two birds with one stone", i.e., sum up the results of the war, and also establish the UN as the first truly global multifunctional organization in history to regulate the entire complex of interstate relations in the post-war world, were largely unjustified. This groundbreaking move in the realm of international relations — not to negotiate a peace treaty, but to establish a post-war system of interstate relations, as if a treaty had really been signed — spared the planet from a direct military confrontation between the Soviet Union and the United States, the two superpowers, for nearly half a century. However, this approach could not be maintained indefinitely.In the new era, characterized by elements such as unipolarity and the collapse of the Soviet Union, as well as the political and economic hegemony of the United States, this approach to adressing the

fundamental political and legal challenges has proven to be ineffective and unproductive. In the absence of significant international competition, the liberal internationalism of the United States at the beginning of the 21st century started to take on features that were initially unexpected: foreign policy radicalism and a greater interest in aggressive expansion. It has increasingly been characterized as a liberal hegemony actively using means such as organizing new coups d'états ("color revolutions"), as well as illegal, that is, not approved by the UN Security Council, trade and economic sanctions. Having recognized the limitations of "soft power", the United States has placed its bets on creating a hierarchical framework for the global political and economic order, with Washington and its closest European allies at the top. The author persistently and convincingly leads the reader to the following conclusion: only the European powers' concentration on their own security, without taking into account the interests of extra-regional players, involving all countries in the negotiation process, respecting the principle of equal and indivisible security and with the goodwill of all parties, can bring back the lost peace to the continent and restore freedom and political subjectivity to Europe.

In the book's Conclusion, R. Sakwa asks: was there an opportunity in the post-Cold War era to manage great power relations to ensure that peace and development took primacy over conflict and hierarchy? The negative answer to this question follows logically from the entire book, which proves that the United States, as the winning power in the Cold War, never seriously considered the prospect of mutually beneficial cooperation and the rejection of its own hegemony. During the Cold War, the main contradiction in the international system was the confrontation between the right of nations to self-determination and the principle of territorial integrity, reflected in the Preamble to the UN Charter. In the new era, another fundamental contradiction has been added: between the principle of indivisible security and the freedom of sovereign states to choose their military and political allies. This leads to the major concern with which the author concludes his study: the conflict in Ukraine may prompt the Western bloc to dismantle the Charter multilateralism altogether. In its place, Washington may try to put up a forum such as The Union of Democracies, presenting it as an alternative to the UN. Then, according to R. Sakwa, the prospect of 'positive peace', as opposed to the mere absence of war, seems "more unrealistic than ever".

В поисках утерянного, но очень нужного мира: зарождение и развитие Второй Холодной войны

Рецензия на книгу: Richard Sakwa (2023). *The Lost Peace: How the West Failed to Prevent a Second Cold War*, Yale University Press: New Haven & London.

Станислав Ткаченко
Доктор экономических наук, профессор кафедры европейских исследований, руководитель магистерской программы «Дипломатия Российской Федерации и зарубежных государств», факультет международных отношений, Санкт-Петербургский государственный университет.
Адрес: Ул. Смольного 1-3, подъезд 8, Санкт-Петербург, 191060, Российская Федерация
E-mail: s.tkachenko@spbu.ru