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 cultures. 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 representation 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 salience," 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.

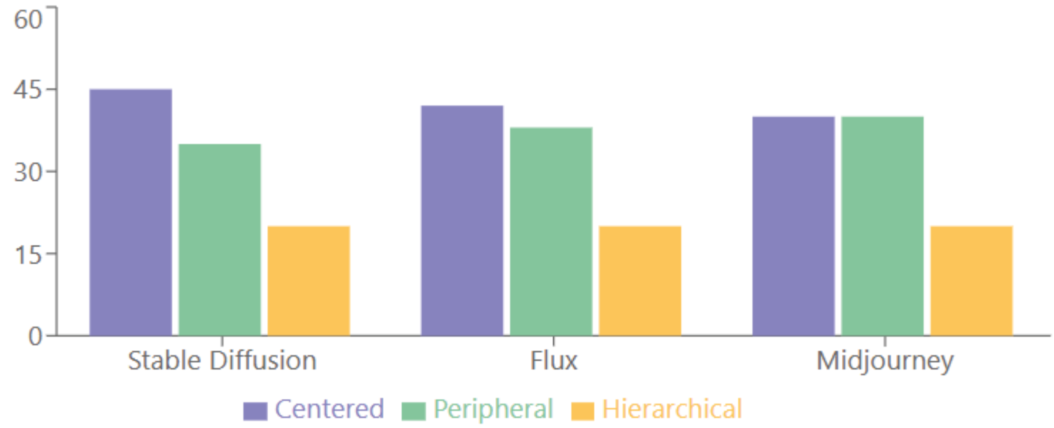

Figure 1. Compositional Patterns Across Platforms

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.

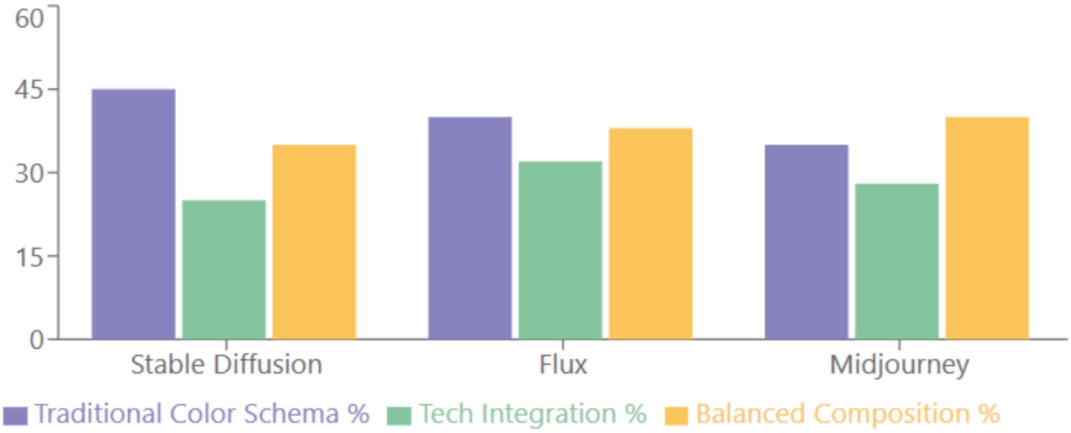

Figure 8. Technical Capabilities Across Platforms

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_CsBawps4Sl1dlFo1tUY/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