The Hype vs. the Reality of Generative AI in Elections: The Case of the 2024 Indian Elections

December 2, 2025

Kiran Garimella
Simon Chauchard

Abstract

The 2024 Indian General Election was defined by pervasive warnings that generative AI would unleash a tsunami of deepfakes and fundamentally destabilize democratic processes. To empirically test these predictions, we analyzed over 5.5 million messages from more than 6,000 WhatsApp groups, obtained through a large-scale data donation project in India. Our findings reveal a profound disconnect between pre-election hype and ground truth: generative AI constituted only 0.7% of the image dataset, with no evidence of a viral "deepfake" epidemic. Furthermore, a qualitative taxonomy of this content reveals that the technology was rarely deployed for deception. Instead, it was primarily utilized for "phatic" labor—automating the production of hyper-stylized "Good Morning" greetings, religious iconography, and patriotic art. We argue that AI adoption was constrained by technological friction and by economic incentives, as traditional "cheapfakes" remained comparatively more cost-effective alternatives for political actors. These results challenge the prevailing focus on the threat of AI-driven messaging, suggesting that the technology may *not* always lead to epistemic disruption, but rather to the efficient reproduction of existing social rituals.

Keywords: Generative AI, WhatsApp, India, Elections, Propaganda.

1 Introduction

The 2024 global election cycle was widely heralded as the "year of generative AI," a turning point where synthetic media would fundamentally alter the mechanics of democracy. In the months preceding India's general election—the world's largest democratic exercise with over 900 million eligible voters—academic experts, security agencies, and technology vendors coalesced around a narrative of "enchanted determinism" (Campolo and Crawford 2020). Predictions ranged from a tsunami of undetectable deepfakes to the automated micro-targeting of propaganda at a scale beyond human comprehension (Cybersecurity and Agency 2024; Elliott 2024). This discourse posited that generative AI acted as an unstoppable, exogenous shock, rendering traditional safeguards of electoral integrity obsolete (Bueno de Mesquita et al. 2023). Accompanying this anxiety was a surge in commercial activity, with political consultancies and startups marketing AI-powered solutions for "personalized voter outreach," creating a feedback loop that reinforced the perception of AI's inevitable omnipresence (Rebelo 2024).

However, despite the highly anxious tone of these predictions, few datasets exist to assess the empirical reality of the 2024 election. While theoretical models of AI harm abound, ecological data from active political environments is scarce. Understanding the actual influence of generative AI is uniquely difficult in the Global South, where political communication occurs primarily on end-to-end encrypted platforms like WhatsApp. Unlike public networks (e.g., X or Facebook) where API access allows for broad surveillance, WhatsApp's dark architecture obscures the provenance and spread of content. Consequently, the existing literature suffers from a "streetlight effect", focusing on visible, public platforms or small-scale qualitative interviews, leaving a profound gap between our anxieties about AI and our knowledge of its ground truth.

This paper addresses that gap by providing the first large-scale, privacy-preserving empirical assessment of generative AI's actual prevalence in the Indian electoral ecosystem. We leverage a unique dataset of over 5.5 million messages collected from 6,000+ WhatsApp groups via a data donation infrastructure deployed across the Northern Indian state of Uttar Pradesh, the largest Indian state and a key political battleground in these elections. To overcome the challenges of detection at scale, we developed a rigorous multi-stage pipeline combining open-weights models (Gemma), state-of-the-art vision language models (GPT-4o), and expert human

annotation. This approach allows us to move beyond simple keyword searches or unverified user reports, distinguishing high-fidelity deepfakes from the benign digital art that characterizes Indian digital culture.

Our findings reveal a stark disconnect between the pre-election hype and the empirical reality. Contrary to predictions of an AI takeover, generative AI content constituted only 0.7% of the total image dataset. Even among highly viral content, where one would expect effective propaganda to cluster, AI-generated media remained a negligible minority compared to traditional, human-generated political messaging. Furthermore, the qualitative nature of this content defied the "disinformation" frame. Rather than malicious deepfakes designed to deceive, the vast majority of AI images were "phatic" in nature: hyper-stylized "Good Morning" greetings, religious iconography, and patriotic digital art.

In addition to providing us with a reality check about the volume and the share of content that was AI-built, these results also reveal what the main driver for the creation of AI content was during this electoral period. Namely, they suggest that the primary driver of AI usage during the 2024 electoral cycle was not the desire to generate more manipulative political content, but rather *efficiency*. Political actors and ordinary users alike appropriated the technology to automate the production of already common, low-stakes social content. In that sense, they fitted AI into the existing aesthetic vernacular of Indian WhatsApp groups instead of using it to create fundamentally different and disruptive types of content. We further identify significant technological friction and economic incentives as limiting factors: for political actors in charge of creating political propaganda ("IT cells"), traditional "cheapfakes" and human-generated rumors remained more cost-effective and persuasive than the computationally expensive and widely distrusted generative AI tools available at the time.

Finally, while generative AI technologies have advanced rapidly since mid-2024, the findings of this study remain critically relevant for the governance of future elections. Critics may argue that the limited impact observed here was merely a function of the technical immaturity of models like DALL-E 3 or Midjourney v6. However, this study establishes a vital sociotechnical baseline: it demonstrates that the constraints on AI disruption are not merely technical, but *sociological* and *economic*. The failure of AI to achieve viral dominance in 2024 highlights the resilience of human distribution networks and the "authenticity paradox," where the glossy

aesthetic of AI triggers user skepticism. As models improve towards photorealism, these human dynamics—trust, distribution, and economic utility—will continue to serve as the primary bottlenecks to adoption. Understanding why the "dog didn't bark" in India in 2024 provides the necessary blueprint for distinguishing between genuine future threats and the persistent noise of enchanted determinism.

2 How Might AI Affect Elections? Theoretical Expectations

The discourse surrounding generative AI in elections has been defined by a sharp tension between theoretical anxiety and empirical ambiguity. This section traces the trajectory of this debate, moving from the "supply-side" predictions of epistemic collapse to the emerging "demand-side" realities of technological friction, economic incentives, and cultural practice in the Global South.

The 2024 global election cycle, dubbed the "super election year", was widely framed through the lens of what Campolo and Crawford term *enchanted determinism*: a discourse where AI is portrayed as an autonomous, unstoppable force that operates beyond human agency (Campolo and Crawford 2020). Security and policy literature heavily amplified this narrative. The Munich Security Conference characterized the period as "AI-pocalypse Now?", predicting that the marginal cost of creating high-fidelity disinformation would approach zero, leading to a tsunami of synthetic propaganda (Carr and Köhler 2024). Scholars argued that the modal richness of AI and its ability to generate photorealistic video and audio, would bypass traditional skepticism, creating a post-truth environment where shared reality is fundamentally fractured (Christopher and Bansal 2024). This supply shock" thesis assumed that because the *capacity* to create deepfakes existed, the *deployment* would be inevitable and ubiquitous (Ferrara 2024).

Countering the determinist narrative, a growing body of literature on the political economy of disinformation suggests that propagators are rational economic actors constrained by technological friction. Paris and Donovan introduced the concept of "cheapfakes"—manipulations achieved through accessible tools like splicing, context-shifting, or speeding up video—arguing that these low-tech methods often yield higher returns on investment than computationally expensive deepfakes (Paris and Donovan 2019). Empirical work supports this cheapfake preference.

Vaccari and Chadwick (2020) and Dobber et al. (2021) found that deepfakes do not necessarily alter voter attitudes more effectively than textual misinformation or simple edits. Furthermore, qualitative research indicates that audiences often view the 'synthetic perfection' of AI-generated content with suspicion, preferring 'analog authenticity' where imperfections serve as markers of truth. Participants expressed concern that unlike traditional photography, the limitless potential of AI to fabricate visuals threatens the concept of an objective, shared reality (Strikovic and Cools 2025). This literature provides a theoretical basis for our finding that political actors in India continued to prioritize human-generated content: the *cost* of upgrading to AI outweighed marginal potential benefits in terms of *persuasion*.

While the direct impact of deepfakes has been limited, scholars identify a secondary, perhaps more pernicious effect: the Liar's Dividend. Schiff, Schiff and Bueno (2025) demonstrate that the mere *public awareness* of generative AI enables political actors to dismiss authentic scandals as fake news or deepfakes. This creates an 'authenticity paradox' where the threat of AI causes more epistemic damage than the AI itself. However, this dividend has a limit. Schiff et al. find that this strategy is highly effective against text but faces resistance against video evidence, suggesting that the public retains a residual trust in visual evidence that 2024-era AI has not yet fully eroded. This dynamic may help explain the low prevalence or the low virality of AI content in social media datasets; as awareness of the technology spreads, the distinctive glossy aesthetic of Midjourney or DALL-E images may now function as a heuristic warning signal—a digital stigma—rather than a tool of deception.

Finally, to understand the prevalence of phatic content such as benign "Good Morning" AI images, we must look to the anthropology of digital communication in India. Scholars of the Global South have long noted that platforms like WhatsApp are not just information conduits but infrastructures for "phatic labor"—the social work of maintaining relationships through small, non-informational gestures (Miller et al. 2016). In the Indian context, the "Good Morning" message is a massive cultural phenomenon, representing a form of digital care that bridges the gap between older, non-English speaking demographics and the internet (Escoe, Martin and Salerno 2025). While elite discourse often dismisses this content as spam, anthropological accounts frame it as a vital social glue. Our findings align with Daepp et al. (2025), who observe that creators in India often use AI not for deception (deepfakes) but for spectacle and efficiency using tools to

quickly generate the hyper-saturated, devotional, or patriotic imagery that fuels this phatic economy. Besides, beyond mere citizens, political actors themselves appear to frequently post such content, likely in a effort to cultivate their audience (Chauchard and Garimella 2022). Following this logic, AI need not be weaponized to attack opponents, and is more likely to be domesticated to serve social media users´ relentless demand for social connection.

While existing literature has successfully mapped the *potential* for AI disruption (supply-side predictions) and the *mechanisms* of its theoretical efficacy (experimental psychology), there remains a critical paucity of ecological data from active political environments. Most current studies rely on small-scale qualitative interviews (Daepp et al. 2025), platform-specific data restricted to the Global North (Chen et al. 2025), or theoretical modeling of threat vectors (Ferrara 2024). Our work bridges this gap by providing the first large-scale, privacy-preserving empirical assessment of generative AI's actual prevalence in the world's largest democracy. By shifting the unit of analysis from *what could happen* (theoretical capability) to *what actually happened* (empirical prevalence), this study offers a necessary corrective to the enchanted deterministic narratives that have dominated the field, re-grounding the discourse in the material realities of friction, cost, and cultural practice.

3 Methods

3.1 Dataset

To empirically assess the prevalence of generative AI content, we overcome the methodological challenges posed by WhatsApp's end-to-end encryption through a large-scale, privacy-preserving data donation infrastructure. Unlike public platforms (e.g., X/Twitter or Facebook) where data can be scraped or accessed via APIs, WhatsApp's architecture necessitates a user-centric collection approach.

In this piece, we thus rely on data from a large WhatsApp data donation program that took place in Uttar Pradesh (India) in 2024. Though we detail our data collection tool and strategy at length in a related piece (Garimella and Chauchard 2025), note that the data collection underwent thorough ethics review and received approval at both [Author 1´s institution] and [Author 2´s institution], and through the funder´s own ethics review process [Funder´s name, grant

number and ethics review reference]. In summary, we recruited a cohort of 3,547 participants across 9 districts of the state of Uttar Pradesh (UP), India. We focused on Uttar Pradesh as it is India's most populous state and a critical political battleground that often serves as a bellwether for national electoral trends. Participants were recruited through on-the-ground partnerships to ensure the sample represented a broad sociodemographic spectrum. The cohort was stratified to capture diversity across key demographic axes, including age, income levels, and caste groups. This stratification is vital for ensuring that our analysis of information consumption captures the experiences of the general electorate rather than being skewed toward tech-savvy or urban elites.

Consenting participants donated data from the large WhatsApp groups - groups of 4 members or more - that they were members of (one-on-one conversations, because they were likely to be of a more private nature, were by default excluded from the donation). We utilized a custom-developed data donation tool designed with a "privacy-by-design" architecture (Garimella and Chauchard 2025). Extensive anonymization was performed locally on the participants' devices before any data was exported to our secure servers. This on-device processing included the removal of personally identifiable information such as phone numbers, names, and email addresses. Furthermore, to protect the privacy of non-consenting group members, faces appearing in images were automatically detected and blurred at the source. This rigorous protocol ensured compliance with ethical standards while allowing us to analyze the content circulating in semi-public digital spaces.

Data collection spanned the critical window of the Indian general election, from March to July 2024. This timeframe captures the pre-election campaigning intensity, the polling phases, and the immediate post-election period. The resulting corpus comprises approximately 6,500 unique WhatsApp groups, yielding a total of 5.5 million messages. Given that generative AI is most frequently discussed in the context of visual misinformation and synthetic media, our analysis specifically isolates the 2.1 million images contained within this dataset.

Another crucial feature of our dataset is the preservation of metadata regarding message dissemination. WhatsApp utilizes specific tags to indicate how content travels across the platform. Messages are flagged as "Forwarded" if they have been passed along once, and "Forwarded many times" if they have been forwarded through a chain of five or more hops. This distinction allows us to differentiate between organic, low-level sharing and content that has

achieved significant virality. By retaining these indicators, we can weight our analysis toward content that actually reached a wide audience, distinguishing between high-visibility political communication and niche or ephemeral content.

3.2 Detecting Generative AI Content

Given the intense discourse surrounding the 2024 election, one might hypothesize a high volume of synthetic media circulating in our dataset. To empirically test this, we focused our detection efforts primarily on image content. This decision is grounded in recent findings that the vast majority (approximately 76%) of AI-generated content on social platforms is image-based rather than video-based (Corsi, Marino and Wong 2024). Furthermore, qualitative studies of political campaigning infrastructure suggest that political IT cells continue to prioritize static imagery, such as meme templates, posters, and greeting cards, over complex video generation, due to ease of production and bandwidth constraints in the Global South (Daepp et al. 2025).

While recent work in the U.S. context has demonstrated that state-of-the-art Vision Language Models like GPT-4o can identify generative AI images with high accuracy using simple zero-shot prompting (Chen et al. 2025), applying this method directly to our entire corpus of over 2 million images was computationally and financially prohibitive. Therefore, we designed a multi-step, cascading detection pipeline designed to maximize recall in the early stages and precision in the later stages.

We began by selecting a random 10% sample of the total image dataset, resulting in approximately 210,000 images. To process this volume efficiently, we employed the open-weights Gemma 3 27B model as a first-pass filter. This model was tasked with flagging any potential AI-generated content. This initial sweep identified 10,500 images as potentially AI-generated.

These flagged images were then subjected to a more rigorous verification step using the GPT-4o model, which currently represents the industry standard for multimodal understanding. This second stage filtered the candidates down to 3,100 images. Finally, to eliminate subtle false positives that automated systems often miss—such as hyper-edited "Good Morning" messages or traditional digital art common in Indian WhatsApp groups—an expert human annotator manually reviewed the remaining 3,100 images. The annotator identified approximately 1,500 images

as generative AI with high confidence.

For the automated stages of the pipeline, we utilized a zero-shot prompt designed to distinguish between synthetic generation and standard digital manipulation. Crucially, the prompt included instructions to account for the privacy-preserving modifications (face blurring) introduced during our data collection process, preventing the models from misclassifying anonymization artifacts as AI "glitches." The prompt used was:

```
Analyze this image and determine if it was created entirely by an AI image
generation model (like DALL-E, Midjourney, or Stable Diffusion). Answer with
a single word first: 'yes' if the image appears to be fully AI-generated,
or 'no' if it appears to be a real photograph, screenshot, text message image,
clipart, poster, meme, or other non-AI-generated content. Sometimes the faces
are blurred due to anonymization so don't be fooled by that.
```

This rigorous three-stage process—moving from open-weights filtering to proprietary model verification to human validation yielded a final prevalence rate of approximately 0.7% within our sample. In and of itself, this figure stands in stark contrast to the enchanted deterministic narratives that predicted an overwhelming flood of AI-generated content.

3.3 Annotation Framework and Protocols

Following the automated detection pipeline, the final set of candidate images underwent a granular manual annotation process. We developed a comprehensive coding instrument designed not only to verify the presence of generative AI but to understand its qualitative characteristics, deceptive potential, and functional utility in political discourse or in other types of communication. A specialized annotation portal was built to enable contextual review: annotators could examine each of the 3,100 pieces of content alongside the 10 messages sent immediately before and after it, preserving the conversational context in which the content appeared.

Unlike traditional content analysis, tasks that rely on crowdsourcing or multiple novice coders to establish inter-coder reliability, detecting modern generative AI content requires a high degree of specialized visual literacy. The artifacts of generative AI, such as inconsistent lighting textures, subtle anatomical errors in hands or eyes, and specific "glossy" aesthetic markers are often imperceptible to the average observer. We deliberated on the use of multiple coders but

determined that without extensive training, variance in annotation would reflect differences in visual acuity rather than ground truth. Consequently, we prioritized validity over broad reliability by utilizing a single, highly trained research assistant with domain expertise in synthetic media forensics and the visual culture of Indian digital spaces. This expert annotator reviewed all 3,100 candidate images, applying a consistent, rigorous standard that a team of novices would not be able to replicate.

The annotation process evaluated each image across four distinct dimensions. The full codebook and questionnaire are detailed in Appendix A.

1. **Generative Confidence (1-5 Likert Scale):** The primary metric assessed the certainty of AI provenance. The scale ranged from 1 ("Definitely not AI-generated") to 5 ("Definitely AI-generated"), with 3 indicating ambiguous cases where provenance was indeterminate. This scale allowed us to isolate high-confidence instances for our final prevalence statistics while retaining ambiguous cases for additional review.

2. **Perceived Authenticity (Layperson Proxy):** Recognizing that the political impact of an image depends on how it is perceived by the electorate, we coded for "deceptive potential." The expert annotator here was asked to estimate whether an average user—one not actively following developments in generative AI—would perceive the image as human-generated. This was particularly relevant for the high volume of "Good Morning" greetings and religious imagery, where the hyper-stylized aesthetic of AI might be mistaken for traditional digital art by a lay audience.

3. **Production Necessity (Counterfactual Analysis):** We assessed the functional utility of the AI tool by asking: *"Could this image have been easily produced without AI?"* This dimension helps distinguish between revolutionary uses of the technology (e.g., creating photorealistic scenes of events that never happened) and efficiency-driven uses (e.g., generating generic images of deities). For instance, while AI was frequently used to generate idols, similar non-AI digital art is already ubiquitous in this ecosystem, suggesting AI served merely as a cost-saving measure rather than a capability booster.

4. **Thematic and Rhetorical Analysis:** Beyond identifying the medium (AI vs. Human), we

systematically categorized the message content to understand its communicative intent. The annotator was tasked with summarizing the central claim or "main idea" of each image. This involved parsing the rhetorical structure of the message, specifically, whether the content was designed to inform, provoke, or solicit specific behaviors (e.g., voting, sharing, or purchasing). We also tracked observable social signals by coding for group reactions, allowing us to determine if specific types of AI-generated content elicited higher engagement or debate within the groups compared to benign or non-AI content.

5. **Epistemic Integrity and Harm Assessment:** To evaluate the potential for AI-driven disinformation, we employed a nuanced coding framework that transcends binary truth claims. We first assessed the *theoretical verifiability* of each image to distinguish factual assertions from subjective political satire. Beyond explicit falsehoods, we coded for *misleading framing*—content that exploits omission or manipulation to deceive without being technically false—and required qualitative justifications for such classifications. Finally, we screened all content for hate speech and toxicity to determine if generative AI was being weaponized against specific communities or remained confined to benign use cases.

This multi-dimensional framework allows us to move beyond simple prevalence counts to construct a taxonomy of *how* and *why* AI is actually being deployed in the wild.

4 Results

Our analysis of the dataset challenges the prevailing narrative of an AI-dominated election. By rigorously filtering the 2.1 million images through our multi-step pipeline, we found that generative AI content constituted only approximately 0.7% of the total image dataset. This low prevalence serves as the foundational context for the detailed breakdown of content types, quality, and virality presented below.

4.1 Prevalence and Quality

Despite the availability of sophisticated tools, the volume of AI-generated content remained negligible. As shown in Figure 1, detecting this content was not a binary task. A significant portion

of images fell into a "grey zone" (ratings 2-3), exhibiting some aesthetic markers of AI—such as hyper-saturated colors or glossy textures—but lacking definitive artifacts. Only 50% of the flagged candidate set received a high-confidence rating (≥ 4) from our expert annotator. From now on, we will only consider these 48% images with a confidence rating ≥ 4 are refered to as "AI posts".

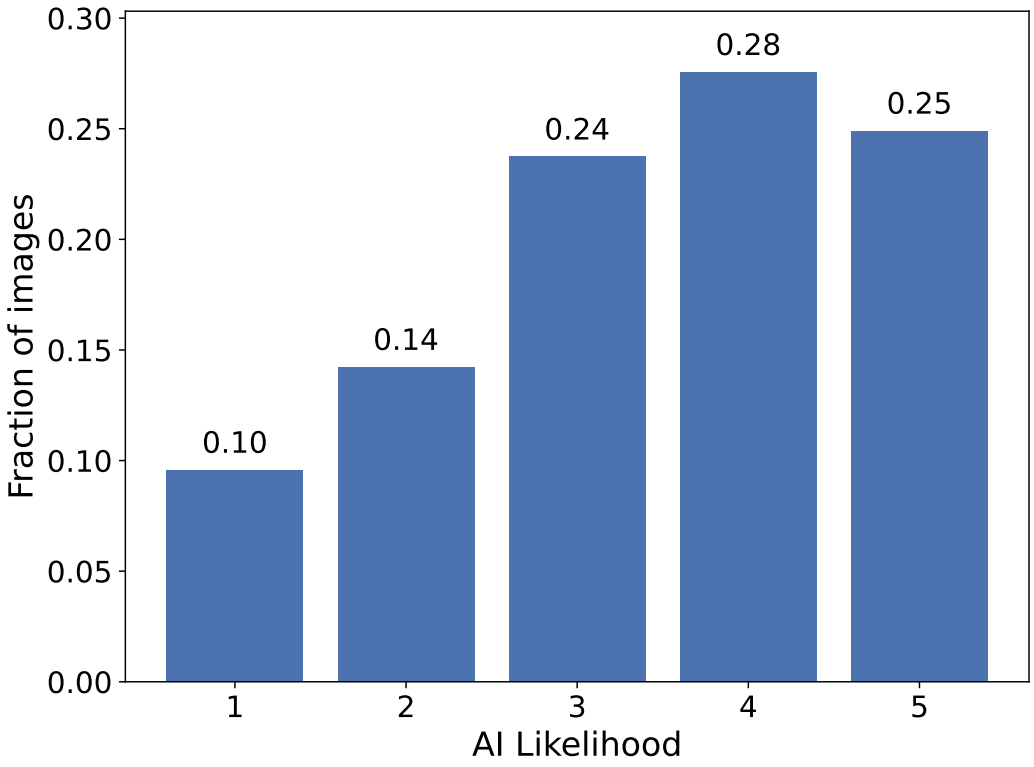


Figure 1: Distribution of manual annotation ratings for AI likelihood (1=Definitely Human, 5=Definitely AI).

Figure A4 (Appendix) presents representative examples across the five-point likelihood scale. A qualitative inspection reveals a striking visual convergence between traditional digital media and AI outputs. Across all rating levels, the images share a consistent vernacular aesthetic (or 'vibe') characterized by religious iconography, floral motifs, and greeting text overlaid on saturated backgrounds. Consequently, high-confidence AI images (Ratings 4-5) are often distinguished not by their subject matter, but by subtle technical artifacts such as incoherent text rendering, anatomical anomalies, or a distinctive hyper-smooth surface texture that deviates from standard digital editing.

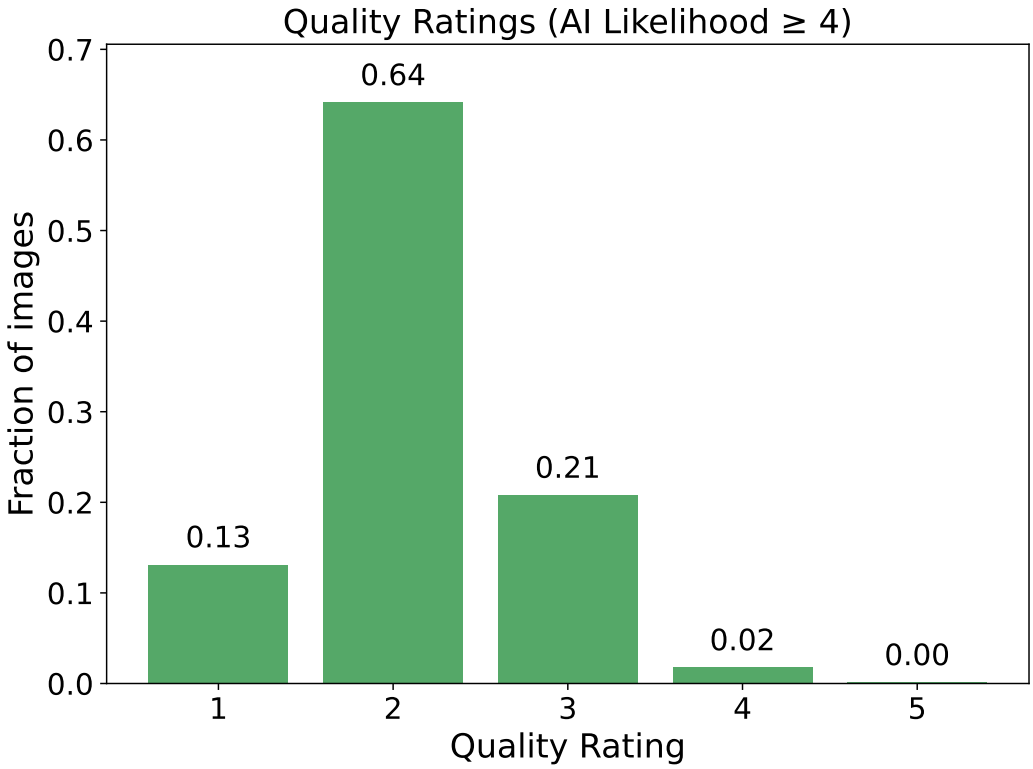


Figure 2: Quality ratings of confirmed AI-generated content (Likelihood ≥ 4).

Furthermore, among the content identified as definitely AI-generated, the technical sophistication was notably low. Figure 2 illustrates the quality ratings of confirmed AI images. The majority of images were rated as "poor quality" (1-2 on the quality scale), characterized by obvious artifacts, anatomical errors, and cartoonish aesthetics. This contradicts the "deepfake" narrative which assumes AI content is indistinguishable from reality; in the Indian context, AI content was often conspicuously artificial.

In sum, these findings challenge the narrative of an AI-saturated information ecosystem. Not only was the overall prevalence of generative AI negligible, but its application was also overwhelmingly banal. Rather than enabling novel forms of manipulation, the technology was primarily co-opted to reproduce existing visual tropes, specifically religious iconography and phatic greetings, serving essentially as a high-tech substitute for traditional digital collages. In the next section, we dig deeper into specific use cases of generative AI content.

4.2 The Taxonomy of AI Content

If AI was not being used to create high-fidelity deepfakes, what was it used for? To answer this, we developed a functional taxonomy based on the primary communicative intent of each image (see Appendix for the full coding instrument). We classified content into distinct categories ranging from social maintenance to political persuasion.

Figure 3 presents the distribution of these categories. The ecosystem is overwhelmingly dominated by what we term "phatic" content (67%). Drawing from anthropological definitions of phatic labor, we define this category as content designed primarily to build or maintain social relationships rather than to convey specific information—this includes the ubiquitous "Good Morning" greetings, festival wishes, and generic devotional imagery.

The second largest category was "Art/Entertainment" (23%), comprising aesthetic digital art, memes, and humorous content devoid of explicit political messaging. In contrast, instrumental categories such as "Logistical" (ads, practical information) or malicious categories such as "Scams" and aggressive political propaganda constituted a minor fraction of the dataset. This distribution confirms that for the average user, generative AI served largely as a tool for social signaling and aesthetic expression rather than information warfare.

A deeper analysis of the phatic category, shown in Figure 4, reveals that approximately one-third of all AI images circulating in our dataset were generic "Good Morning" messages or greetings. This suggests that the primary driver of AI adoption in this election cycle was not deception, but *efficiency*. Users and content creators utilized AI tools to generate quick, visually appealing greeting cards and inspirational quotes, as they presumably had done before the electoral campaign (Chauchard and Garimella 2022), rather than to manipulate public opinion. While we detect very little content that can credibly be classified as fulfilling this goal, we note that this number would in all likelihood have been even smaller before and after the end of the campaign, as interest in politics waned.

Table A1 (Appendix) further corroborates this, showing that high-confidence AI images (Score 5) were disproportionately concentrated in the "Phatic" (73.3%) and "Art" (31.8%) categories, whereas lower-confidence images were often logistical or advertising spam. Visual examples of this benign content such as AI-generated flowers or generic greetings can be seen in

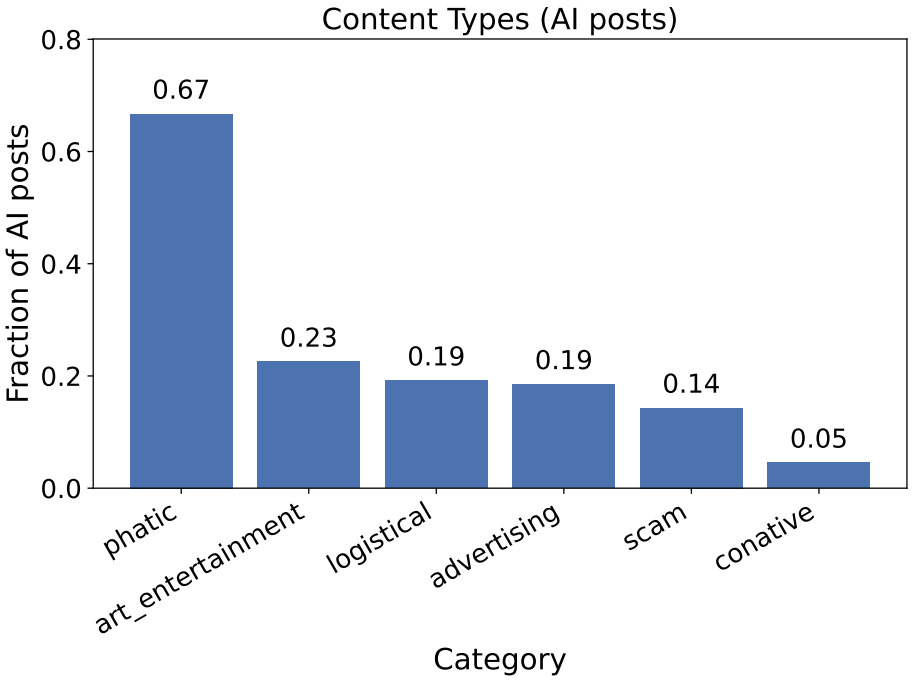


Figure 3: Distribution of AI content by category.

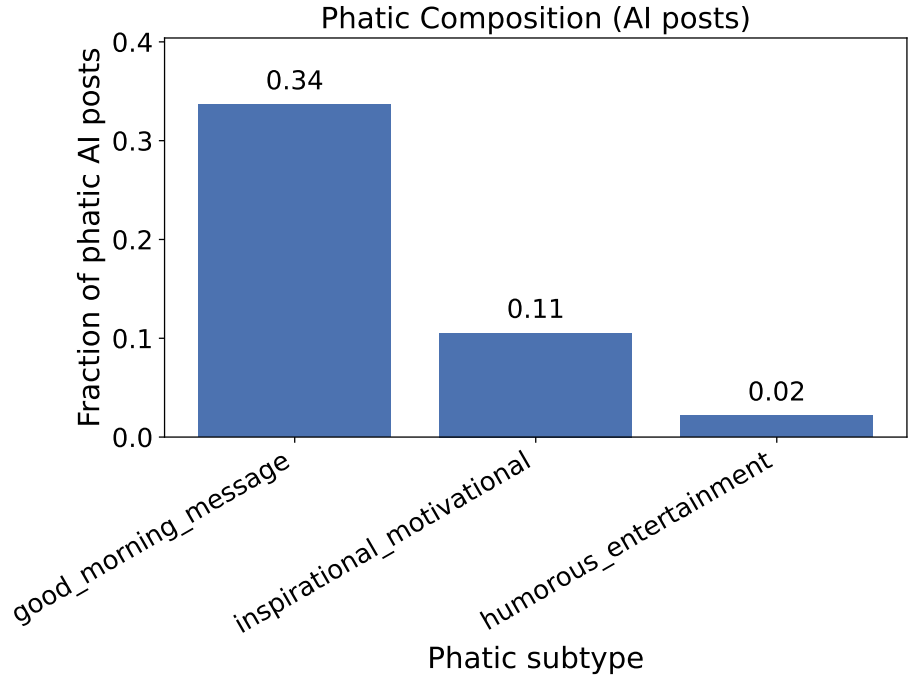


Figure 4: Breakdown of phatic AI content, dominated by greetings and motivational quotes.

Figure A6 in the Appendix.

4.3 Political and Social Signals

While the overall volume of political AI content was low, its distribution reveals significant asymmetries. To distinguish between general cultural content and explicit electioneering, our coding scheme (detailed in the Appendix) differentiated between three categories of relevance. We defined "Socially Relevant" content as images pertaining to general associational life, such as festivals, cultural symbols, or deities, without necessarily invoking political actors. In contrast, the "Political" tag was reserved for content discussing governance or public affairs, while "Partisan" was strictly applied to messages explicitly promoting or attacking a specific party or leader.

As shown in Figure 5, the results are stark. Over half (56%) of the AI content was broadly "Socially Relevant," reflecting the technology's use in cultural signaling. Conversely, only 4% was explicitly "Political," and a mere 3% met the strict criteria for "Partisan" propaganda.

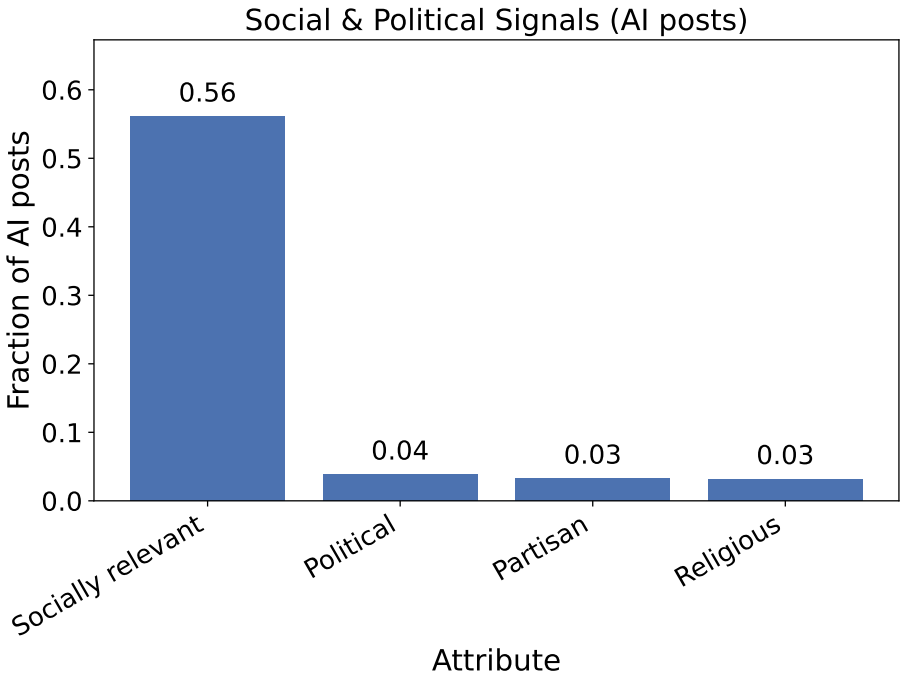


Figure 5: Proportion of AI content containing social, political, or partisan signals.

We also examined the "habitats" of this content to understand the contexts in which it circulated. Figure A3 (Appendix) reveals that nearly 50% of AI images appeared in Religious

groups, followed closely by "Political" (31%) and "Hindutva/BJP" (25%) groups. This distribution highlights a significant functional dynamic: while the AI content itself is rarely explicitly political (as shown in Figure 5), it is heavily utilized within political and partisan infrastructures. This suggests that benign AI-generated imagery, such as religious idols and morning greetings, plays a critical role in the maintenance of ties and in sociality between group members. Prior research had shown that in these political spaces, high-aesthetic, low-stakes content likely serves to keep groups active and socially lubricated, maintaining user engagement and communal bonding during lull periods between explicit political mobilization (Chauchard and Garimella 2022). We show here that this extends to campaign periods during which more openly political, or even partisan, content would be expect to dominate, as well as to the specifically AI-built subset of the content. Despite the proximity of an election happening in a context of extreme polarization in India, users of these threads had recourse to AI technology first and foremost to salute one another and send wishes.

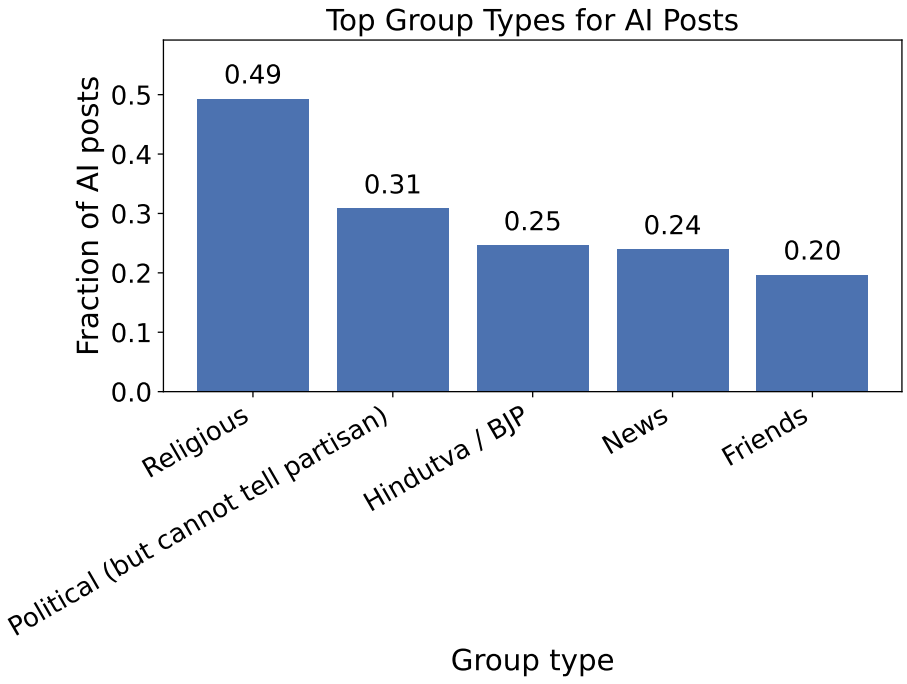


Figure 6: Distribution of AI content across different WhatsApp group types.

4.4 The "Dog That Didn't Bark": Misinformation and Hate

This relative absence of political content within the AI-build subset of our data drives another surprising result. Contrary to predictions of deepfake-driven chaos, misinformation and hate speech were indeed almost non-existent in our AI subset. Hate speech constituted approximately 1% of the AI images.

A qualitative analysis of the rare misinformation examples (Figure 7) reveals that they seldom constituted malicious political fabrications in the conventional sense. Instead, they were primarily instances of "myth-making," characterized by pseudoscientific health claims (e.g., the "Sanjeevani Mantra" (translation: Life-restoring chant) for longevity), exaggerated religious miracles, or hyper-idealized, futuristic renderings of public infrastructure. Similarly, the sparse instances of hate speech (Figure 8) were technically crude and relied heavily on established sectarian tropes, specifically narratives regarding "Love Jihad" or demographic anxiety, rather than novel, AI-generated falsehoods. Consequently, the technology was not deployed to engineer new forms of deception, but rather to visually amplify existing folk theories and social prejudices.

((a)) ((b)) ((c))

Figure 7: Misinformation examples

((a)) ((b)) ((c))

Figure 8: Hateful examples

5 Discussion

Our investigation into the 2024 Indian General Election offers a corrective to the "enchanted determinism" that characterized pre-election discourse (Campolo and Crawford 2020). The dominant narrative, propelled by security agencies, media outlets, and technology vendors, posited that generative AI would act as an unstoppable, exogenous shock to the democratic system, flooding the zone with undetectable deepfakes and personalized propaganda.

The empirical reality is different. With generative AI constituting merely 0.7% of the image dataset, the anticipated "tsunami" never materialized. Even when generative AI was used, it was just to complement existing workflows, e.g. to create memes and phatic content. This disconnect forces us to re-evaluate the incentives and constraints governing political communication in the Global South. The failure of AI to transform the 2024 election was not merely a matter of technological immaturity. The practices of political parties (such as the ruling BJP) in terms of content creation are likely inelastic in the short term — costs and skills acquisition constitute significant hurdles. Besides, there is no reason to expect that the apparition of a new technology would change the preferences of users in terms of content. Insofar as phatic, rather than political

or partisan content, has dominated and continues to dominate social media posting in India, it would have been surprising to see *political* deepfakes take over. Logically enough, AI was used, however disproportionally so by creators of phatic content.

5.1 The Friction of Reality: Why Predictions Failed

The gap between the predicted "AI Election" and the ground truth can be attributed to what we term *technological friction*. The enchanted view of AI assumes that because a technology *can* be used for deception, it *will* be. However, this ignores the logistical realities of political actors that would most likely be in charge of producing such content in India, the famous "IT cells" of parties, for instance (Mehta 2024). For a political operative, the goal is not to use the most advanced technology, but to use the most effective one. Creating a high-fidelity deepfake that survives scrutiny requires significant prompt engineering, consistency checks, and often post-processing is a high-friction workflow. In contrast, "cheapfakes"—miscontextualized videos, edited newspaper clippings, or traditional text-based rumors—are low-friction, cheap to produce, and historically proven to drive engagement (Dan et al. 2021). Our findings suggest that AI had, as of 2024 and in India, not yet crossed the threshold of utility after which it is more efficient than a cheapfake for the specific purpose of negative campaigning. The "dog that didn't bark" in this election suggests that disinformation actors are rational economic agents who prioritize return on investment over technological novelty. For instance, see Figure A5 shows examples of misinformation images which were not generated using generative AI. We can clearly see most examples are simple images with text on a plain background or on an image, or image collages, or simple animations.

It is crucial, however, to contextualize these findings from the technological vantage point of 2024. The models available during our data collection window (e.g., Midjourney v6, DALL-E 3) still struggled with text rendering and specific Indian cultural markers. While the technology has and will undoubtedly improve, our data may suggest that the social demand for high-fidelity deception is lower than assumed; in a context of high social and partisan polarization, human-generated misinformation may suffice.

Finally, it is imperative to acknowledge that the disparity between prediction and real-

ity does not necessarily invalidate the concerns raised by the expert community. Indeed, the high-volume discourse surrounding the "AI Election" may have functioned as a self-defeating prophecy in the most positive sense. The widespread media coverage and pre-bunking campaigns led by civil society organizations likely increased digital literacy, priming the electorate to be hyper-vigilant. By creating a climate of suspicion around synthetic media, these predictions may have inadvertently raised the bar for successful deception, rendering low-quality deepfakes ineffective before they could proliferate.

5.2 Supply-Side Dynamics: Automation, Not Deception

If AI was not used to deceive, what was it used for? Our finding that the majority of AI content falls into "Phatic" and "Religious" categories reveals that Indian users have adopted AI primarily as a *cost-saving mechanism* to keep producing pre-existing contents in a more efficient manner. In the high-volume environment of Indian WhatsApp, the exchange of "Good Morning" messages and religious greetings is a vital ritual for maintaining social ties. Generative AI, with its tendency toward hyper-saturated colors, symmetrical composition, and idealized imagery, fits the existing aesthetic vernacular of this genre perfectly. AI was merely used to automate the production of the "glossy" aesthetic that Indian users already preferred.

Furthermore, our finding that 84% of partisan AI content (a small share of all AI content, as shown above) favored the ruling BJP challenges the "democratization" hypothesis (Kreps and Kriner 2023). Early optimists argued that AI would level the playing field, allowing resource-poor parties to generate high-quality content at near-zero cost. Instead, we observe that AI adoption mirrors existing offline power dynamics. The BJP's dominance suggests that the bottleneck in digital campaigning is not *content production* (which AI solves), but *distribution networks* (which AI does not). The ability to push content to millions of voters still requires the massive human infrastructure and WhatsApp group networks that the ruling party has spent a decade cultivating (Shih 2023). AI simply allowed the strong to become more efficient, rather than empowering the weak.

5.3 Demand-Side Dynamics: The Heuristics of Skepticism

The limited virality of AI content (Figure A1, Appendix) provides critical insight into audience reception. The fact that AI images had lower viral peaks than human content suggests that users may possess an intuitive skepticism (or perhaps indifference) toward the "synthetic" aesthetic. We hypothesize that the specific visual artifacts of 2024-era AI (plastic skin textures, incoherent background details) act as a heuristic signal for "low-effort" or "spam" content. Just as email users learned to ignore messages with poor grammar, WhatsApp users may be learning to filter out the glossy, uncanny look of AI images. Far from being the "perfect weapon" of influence, overt AI generation may currently impose a penalty on credibility. The content is recognized not as documentation of reality, but as "digital art" or "forwarded clutter," reducing its potential to mobilize political anger or belief.

5.4 Implications for Democratic Resilience

These findings have profound implications for AI governance. Policy responses globally have focused on technical solutions like watermarking or detection to prevent deepfakes. However, our data suggests that policymakers may be fighting a theoretical war while ignoring the empirical one.

Crucially, the relevance of these findings extends beyond the specific technical capabilities of the 2024-era models used in this study. Critics might argue that the limited impact we observed was merely a function of the "uncanny valley" artifacts present in tools like DALL-E 3 or Midjourney v6. However, this technological determinism ignores the enduring sociological and economic constraints we identified. The friction against AI adoption was not just *fidelity*, but *utility*. Even as future models achieve perfect photorealism, the economic logic of the "cheapfake"—which creates high engagement with zero compute costs and minimal effort—remains a formidable competitor. Our study establishes a baseline demonstrating that without a shift in the *demand* for high-fidelity deception, improvements in the *supply* of AI capabilities may yield diminishing returns in actual electoral influence.

First, the obsession with "deepfakes" distracts from the ongoing prevalence of "shallow" misinformation and hate speech, which constituted the vast majority of problematic content in

our dataset. Second, the real danger of AI in elections may not be the *quality* of the lie, but the *quantity* of the noise. By automating the creation of phatic and religious content, AI allows actors to flood information channels with benign noise, potentially drowning out genuine political discourse and exhausting user attention.

Finally, the enchanted deterministic narrative itself may pose a risk to democratic integrity. By continuously warning voters about "undetectable" AI, experts may inadvertently contribute to a "liar's dividend," where political actors can dismiss genuine incriminating evidence as "AI-generated." Our study suggests that to preserve electoral integrity, we must pivot from fearing a hypothetical technology to understanding the mundane, efficient, and culturally specific ways it is actually being adopted.

References

Bueno de Mesquita, E., B. Canes-Wrone, A. B. Hall, K. Lum, G. J. Martin and Y. R. Velez. 2023. Preparing for Generative AI in the 2024 Election: Recommendations and Best Practices Based on Academic Research. Technical report Stanford Graduate School of Business & University of Chicago Harris School of Public Policy. Available at https://www.gsb.stanford.edu/faculty-research/publications/preparing-generative-ai-2024-election-recommendations-best-practices.

Campolo, A. and K. Crawford. 2020. "Enchanted Determinism: Power without Responsibility in Artificial Intelligence." *Engaging Science, Technology, and Society* 6:1–19.
URL: *https://doi.org/10.17351/ests2020.277*

Carr, Robert and Paul Köhler. 2024. Ai-pocalypse now? disinformation, AI, and the super election year. Munich Security Conference. https://doi. org/10.47342/VPRS3682.

Chauchard, Simon and Kiran Garimella. 2022. "What circulates on partisan WhatsApp in India? Insights from an unusual dataset." *Journal of Quantitative Description: Digital Media* 2.

Chen, Zhiyi, Jinyi Ye, Beverlyn Tsai, Emilio Ferrara and Luca Luceri. 2025. Synthetic politics: Prevalence, spreaders, and emotional reception of AI-generated political images on X. In *Proceedings of the 36th ACM Conference on Hypertext and Social Media*. pp. 11–21.

Christopher, N. and V. Bansal. 2024. "Indian voters are being bombarded with millions of deepfakes. Political candidates approve." *WIRED* . Available at https://www.wired.com/story/indian-elections-ai-deepfakes/.

Corsi, Giulio, Bill Marino and Willow Wong. 2024. "The spread of synthetic media on X." *Harvard Kennedy School Misinformation Review* 5(3):1–19.

Cybersecurity and Infrastructure Security Agency. 2024. "Risk in Focus: Generative AI and the 2024 Election Cycle." Online. Available at https://www.cisa.gov/resources-tools/resources/risk-focus-generative-ai-and-2024-election-cycle.

Daepp, Madeleine IG, Alejandro Cuevas, Robert Osazuwa Ness, Vickie Yu-Ping Wang, Bharat Kumar Nayak, Dibyendu Mishra, Ti-Chung Cheng, Shaily Desai and Joyojeet Pal. 2025. "Generative Propaganda." *arXiv preprint arXiv:2509.19147* .

Dan, Viorela, Britt Paris, Joan Donovan, Michael Hameleers, Jon Roozenbeek, Sander van der Linden and Christian von Sikorski. 2021. "Visual mis-and disinformation, social media, and democracy." *Journalism & Mass Communication Quarterly* 98(3):641–664.

Dobber, Tom, Nadia Metoui, Damian Trilling, Natali Helberger and Claes De Vreese. 2021. "Do (microtargeted) deepfakes have real effects on political attitudes?" *The International Journal of Press/Politics* 26(1):69–91.

Elliott, V. 2024. "2024 is the Year of the Generative AI Election." *WIRED* . Available at https://www.wired.com/story/2024-is-the-year-of-generative-ai-elections/.

Escoe, Brianna, Nathanael S Martin and Anthony Salerno. 2025. "That's so cringeworthy! Understanding what cringe is and why we want to share it." *Journal of Marketing Research* 62(4):664–683.

Ferrara, Emilio. 2024. "Charting the landscape of nefarious uses of generative artificial intelligence for online election interference." *arXiv preprint arXiv:2406.01862* .

Garimella, Kiran and Simon Chauchard. 2025. "WhatsApp explorer: A data donation tool to facilitate research on WhatsApp." *Mobile Media & Communication* 13(3):481–503.

Kreps, Sarah and Doug Kriner. 2023. "How AI threatens democracy." *Journal of Democracy* 34(4):122–131.

Mehta, Nalin. 2024. How the Bjp Became the World's Largest Political Party: Organisational Restructuring and the Use of Digital Technologies. In *The New BJP*. Routledge pp. 168–217.

Miller, Daniel, Jolynna Sinanan, Xinyuan Wang, Tom McDonald, Nell Haynes, Elisabetta Costa, Juliano Spyer, Shriram Venkatraman and Razvan Nicolescu. 2016. *How the world changed social media*. UCL press.

Paris, Britt and Joan Donovan. 2019. "Deepfakes and cheap fakes." *Data & Society* 1.

Rebelo, Karen. 2024. "India's generative AI election pilot shows artificial intelligence in campaigns is here to stay." *Center for Media Engagement at The University of Texas at Austin* pp. 12–16.

Schiff, Kaylyn Jackson, Daniel S Schiff and Natália S Bueno. 2025. "The liar's dividend: can politicians claim misinformation to evade accountability?" *American Political Science Review* 119(1):71–90.

Shih, Gerry. 2023. "Inside Hindu Nationalists' Vast Digital Campaign to Inflame India." *The Washington Post* . News article.
URL: *https://www.washingtonpost.com/world/2023/09/26/hindu-nationalist-social-media-hate-campaign/*

Strikovic, Edina and Hannes Cools. 2025. "Reality Re-Imag(in)ed: Mapping Publics' Perceptions and Evaluation of AI-generated Images in News Contexts." *Digital Journalism* .

Vaccari, Cristian and Andrew Chadwick. 2020. "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news." *Social media+ society* 6(1):2056305120903408.

Appendix:
The Hype vs. the Reality of Generative AI in Elections: The Case of the 2024 Indian Elections.

Contents

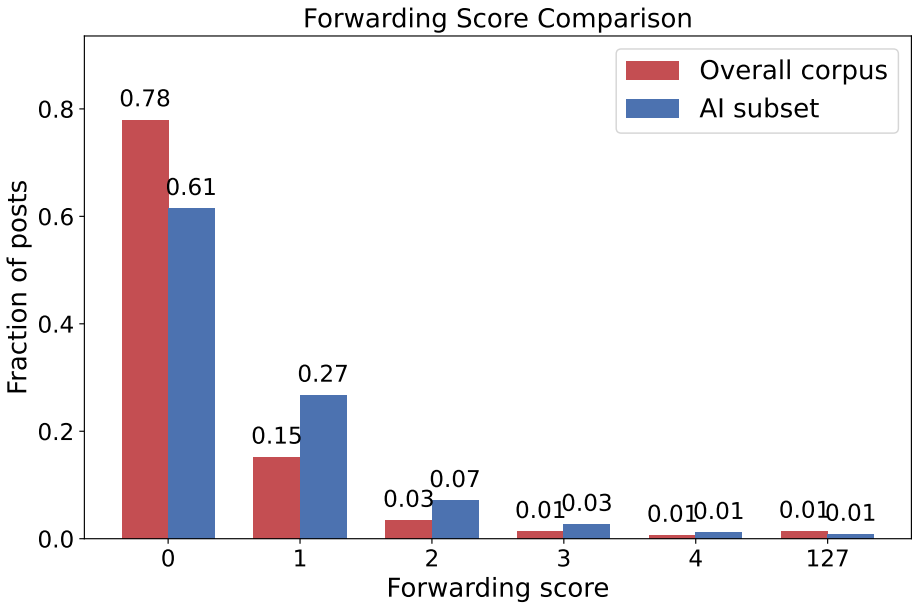
A Additional results

A.1 Virality and Temporal Dynamics

A key question is whether the "novelty" of AI content confers a viral advantage. Figure A1 compares the forwarding depth of AI content against all other images in our dataset, revealing a nuanced distinction between *shareability* and *virality*. Contrary to the assumption that AI content would inherently outperform human content, we find that the advantage is limited to local circulation.

AI-generated images are significantly more likely to be forwarded at least once compared to the general corpus (which is heavily skewed toward unshared content, with 78% having a score of 0 versus 61% for AI). Consequently, AI content shows consistently higher prevalence in short forwarding chains (scores 1, 2, and 3), suggesting that the novelty or aesthetic appeal of these images is effective at generating initial engagement within immediate social circles. However, this advantage dissipates at the threshold of mass virality; for highly forwarded content (marked as viral/127), AI images perform indistinguishably from human-generated material. This indicates that while AI is effective at avoiding obscurity, it does not possess an inherent "super-viral" quality that allows it to dominate the long-tail of distribution.

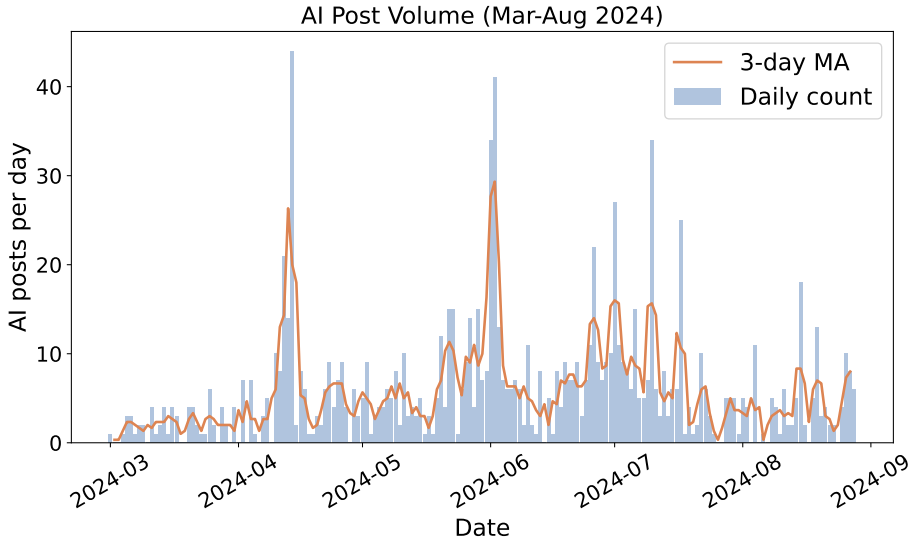
Figure A1: Comparison of forwarding scores (virality) between the AI subset and the overall corpus.



A.2 Temporal Trends

Figure A2 plots the temporal volume of AI posts. While overall volume remained low, we observed distinct spikes on April 19 and June 1. These dates coincide exactly with the first and last phases of the Indian General Election. While this correlation suggests some responsiveness to the electoral calendar, the overall low volume implies these were likely bursts of "Get out the vote" imagery or celebratory content rather than coordinated disinformation campaigns.

Figure A2: Daily volume of AI posts, showing spikes corresponding to election phases (April 19 and June 1).



A.3 Type of AI content by AI Likelihood

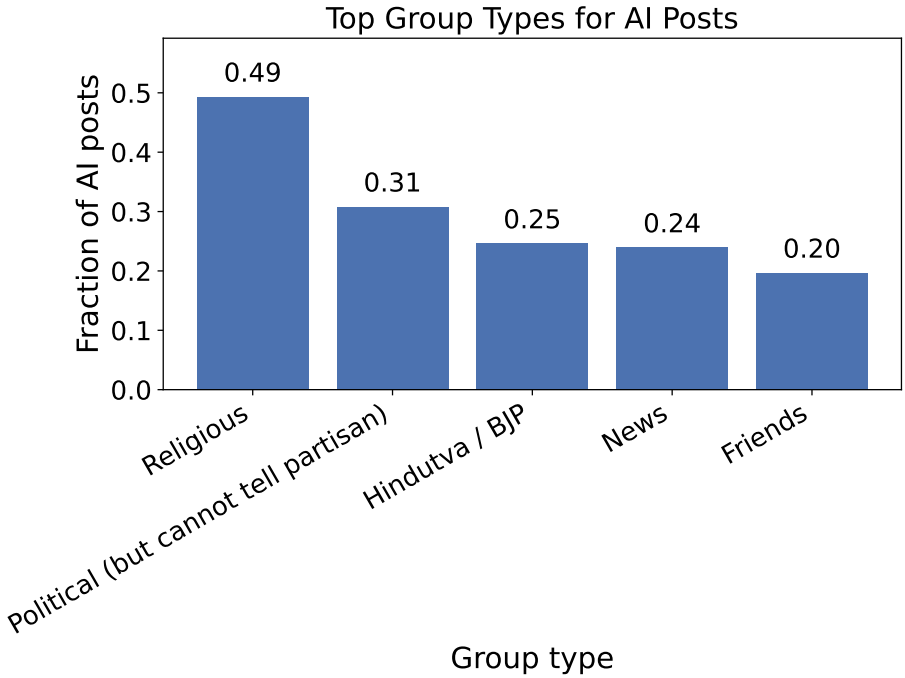
Table A1: Content type share (%) by self AI likelihood.

AI Likelihood	Total	phatic	art_entertainment	logistical	advertising	scam
1	278	55.4	8.6	14.0	11.9	5.4
2	413	81.6	10.2	9.4	8.2	2.4
3	689	64.3	11.5	14.8	13.6	9.4
4	800	60.8	14.1	29.1	28.2	24.0
5	723	73.3	31.8	8.2	7.7	3.6

A.4 Group Types

Types of groups: Figure A3 shows the top groups (note that an image can be shared in multiple groups, so the bars do not add up to 100). Among the AI content, around 50% of them were shared in religious groups and other political groups. Though the religious groups may not be surprising given the composition of the generative AI content, the presence of non political generative AI in political spaces indicates that these groups share such phatic content (Chauchard and Garimella, 2022).

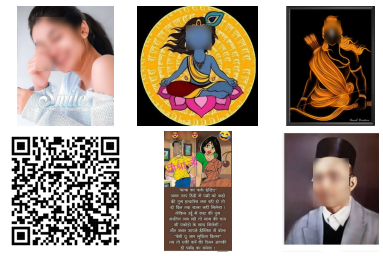
Figure A3: Group types



A.5 Likelihood of AI examples

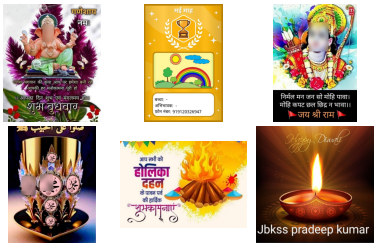
Figure A4 shows examples of images by the annotated likelihood of being AI generated. In our paper, we consider images with a likelihood ≥ 4 as AI generated.

Figure A4: AI Likelihood rating examples

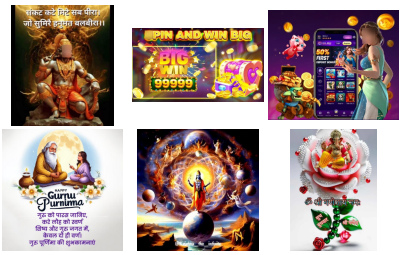


(a) Likelihood=1

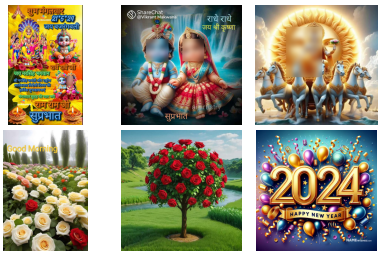
(b) Likelihood=2



(c) Likelihood=3



(d) Likelihood=4

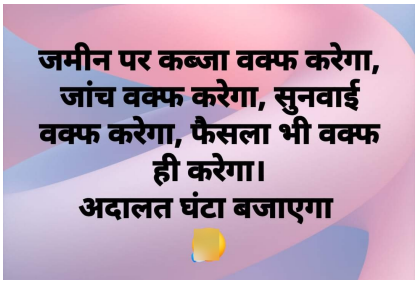


(e) Likelihood=5

A.6 Non-AI generated political propaganda examples

Figure A5 shows examples of non-AI generated political propaganda.

Figure A5: Examples of Non-AI generated misinformation. We can clearly see that most content is simple text on an image, or a collage of images, or simple animations.



(a)

(b)

(c)

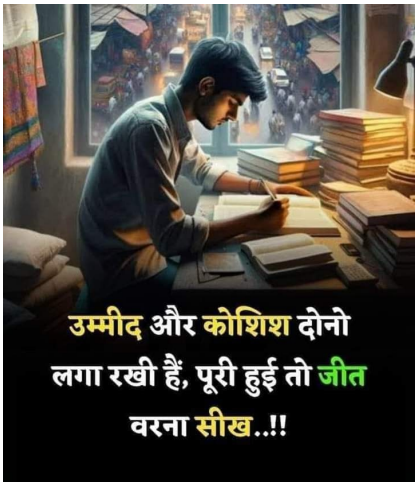
(d)

(e)

A.7 Phatic content examples

Figure A6 shows examples of phatic content.

Figure A6: Examples of Phatic AI Content



(a)

(b)

(c)

A.8 Likelihood of social/political Attribute Share by AI Likelihood

Table A2 shows the likelihood of social and political attribute share by AI likelihood.

Table A2: Social/political attribute share (%) by annotated AI likelihood.

AI Likelihood	Total	Socially relevant	Political	Partisan	Religious
1	278	56.1	4.3	3.6	47.8
2	413	71.9	2.4	1.9	67.6
3	689	51.4	3.3	2.6	48.5
4	800	47.1	1.6	1.4	46.0
5	723	66.0	6.2	5.3	64.0

B Appendix: Annotation Questionnaire

The following questionnaire was used by expert annotators to code the dataset. The annotation process was divided into three main sections: AI Assessment, Socially Contested Assertions about the World, Coding, and Thematic Tagging.

B.1 Section 1: AI Assessment

1. **If you did not know this was AI, to what extent would you think this was made by AI?**
   Scale: 1 to 5
   (1 = Not AI generated, 5 = Definitely AI)

2. **To what extent would someone else (less literate about AI), think it was made by AI?**
   Scale: 1 to 5
   (1 = Not AI generated, 5 = Definitely AI)

3. **How would you rate the quality of AI?**
   Scale: 1 to 5
   (1 = Poor quality, 5 = High quality)

4. **To what extent could this image have been produced without AI?**
   Scale: 1 to 5
   (1 = Could have been made without using any AI, 5 = Without AI, this image cannot exist)

B.2 Section 2: Socially Contested Assertion about the World (SCAW) Coding (If SCAW is present)

1. **Does this message contain a SCAW?**
   - Yes
   - No

2. **What is the claim about?**
   (Summarize the main idea or goal of the claim in the message)

3. **Does some part of this message contain misinformation according to you?**
   - Yes
   - Maybe
   - No

4. **Is this closer to being misinformation?**
   - Yes
   - No

5. **Can the part you think is misinformation be theoretically verified?**

- Yes
- No

6. **Do you find evidence to verify?**
   - Yes
   - No
7. **Please provide a justification / evidence:**
   (Open text input)
8. **Is the claim misleading?**
   (Is the message structured in a way to mislead? e.g., omits details, manipulative framing, etc.)
   - Yes
   - No
9. **Why is the claim misleading?**
   (50–300 characters)
10. **Do people react to this message in a group?**
    - Yes
    - No
11. **Nature of the response:**
    (e.g., Neutral, Positive, Negative)
12. **Associated Emotions (Select all that apply):**
    - Happiness
    - Pride
    - Anger
    - Sadness
    - Shame
    - Fear
    - Disgust
    - Resentment
    - Indifferent
    - Hatred
    - NA

B.3 Section 3: Thematic Tags

Choose the most appropriate tags (where applicable):

1. **Hateful Content**
    - Does the message spread hate against individuals or groups?
    - *If yes, please specify hateful content:* (Text input)
2. **Anti-Muslim**
    - Common theories:
        - Population jihad
        - Love jihad
        - Muslim appeasement
        - Forced conversions
        - Other
3. **Political Content**
    - Is the message related to political topics or parties?
    - **Partisan:**
        - **Partisan for (promotes):** Does the message promote a party, leader, or ideology?
        - **Partisan against (critiques):** Does the message critique or attack a political leader or party?
4. **Religious**
    - Is it connected to religious aspects or iconography?
    - Religious iconography (Checkbox)
5. **Caste** (Checkbox)
6. **Location Scope:**
    - Is the claim about local, national, or international events?

B.4 Section 4: Nature and Intent Characteristics

Does the message also show these characteristics?

1. **Message asking people to do something/alter their behavior**
   (A statement intended to influence behavior)
    - Political behavior
    - Health behavior
    - Everyday/Mundane behavior

2. **Message related to organizing, planning**
   (Related to logistics or organization)

   - Advertising
   - Practical information

3. **Messages meant to build/maintain social relationships (Phatic)**
   (Small talk, humor, greetings etc. shared to connect with others casually)

   - Humorous/Entertainment
   - Good morning message
   - Inspirational/Motivational

References

Chauchard, Simon and Kiran Garimella. 2022. "What circulates on partisan WhatsApp in India? Insights from an unusual dataset." *Journal of Quantitative Description: Digital Media* 2.