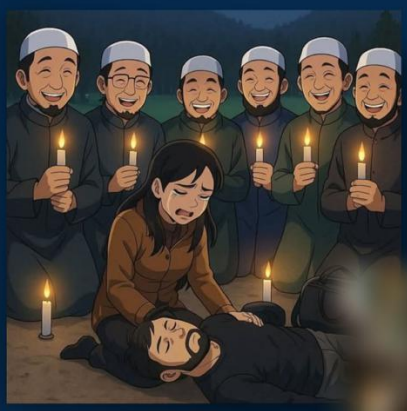
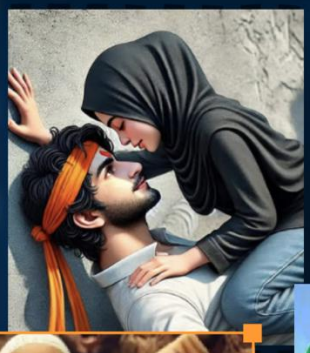
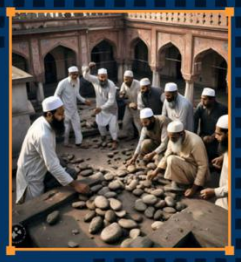


Image Processed



AI-Generated Imagery and the New Frontier of Islamophobia in India

CSOH
Center for the Study
of Organized Hate



CSOH

1629 K St. NW, Suite 300
Washington, D.C. 20006

The Center for the Study of Organized Hate (CSOH) is a nonprofit, nonpartisan think tank based in Washington, D.C. CSOH is strongly driven by its mission to advance research and inform policies that combat hate, violence, extremism, radicalism, and disinformation.

Our research, strategic partnerships, and community engagement programs are guided by the vision of a more inclusive and resilient society against all forms of hate and extremism.

Authors
Nabiya Khan, Researcher, CSOH
Aishik Saha, Researcher, CSOH
Zenith Khan, Data Analyst, CSOH

We thank Apar Gupta, Co-founder and Founder Director of the Internet Freedom Foundation (IFF), for providing the legal perspective that informed our recommendations.

Connect With Us
Website: www.csohate.org
LinkedIn: https://www.linkedin.com/company/csohate/
Instagram: www.instagram.com/csohate
Facebook: www.facebook.com/csohate
X: www.x.com/csohate
TikTok: https://www.tiktok.com/@csohate

For inquiries, please email: press@csohate.org

TABLE OF CONTENTS

1. INTRODUCTION

This report documents the use of generative Artificial Intelligence (AI)[1] to produce and disseminate anti-Muslim visual hate content in India. The use of AI-generated images exploded globally in 2022[2], drawing wide interest and enthusiasm for their many potential uses. However, their role in shaping online hate speech remains underexplored. This is especially true for India's volatile digital and political landscape. While deployment of social media for spreading hate in the Indian context has received significant attention and has been flagged through numerous studies, reports, and journalistic articles, including those produced by the Center for the Study of Organized Hate (CSOH)[3], the deployment of harmful AI-generated content has not yet been analyzed. As ChatGPT seeks to access the potentially vast Indian market by offering an inexpensive subscription plan[4] priced at less than $5 a month, the need for such analysis becomes especially paramount. This report offers an early and urgent intervention into the ways AI tools such as Midjourney, Stable Diffusion, and DALL·E are being used to generate synthetic images that fuel hate and disinformation against India's Muslims. By examining AI-generated images that dehumanize, sexualize, criminalize, or incite violence against Muslims, the report documents how emerging technologies are being weaponized.

The dataset analyzed for this report comprises 1,326 publicly available AI-generated images and videos retrieved from 297 public accounts in Hindi and English across X (formerly Twitter), Facebook, and Instagram between May 2023 and May 2025. Within this time frame, the posts are concentrated from January 2024 to April 2025. The use of AI-generated content focused on spreading hate spikes after January 2024. These 297 public accounts were first selected through purposive sampling, as detailed in the methodology section below. They reflect extremist right-wing voices in the Indian internet space, with an active and consistent record of posting hateful content online, significant follower counts, and high engagement for their posts. The 1,326 posts were identified from the 297 accounts through qualitative coding. Each post was manually verified. The archive was then analyzed for thematic and narrative patterns through coding, through which key categories, subcategories, and tropes of anti-Muslim hateful content were identified. The analysis revealed four main categories across the archive: the sexualization of Muslim women, exclusionary and dehumanizing rhetoric, conspiratorial narratives, and the aestheticization of violence. The report also tracked how such posts were amplified across platforms by Hindu right-wing media outlets and networks.

We have focused on the impact of the content produced by this group of accounts over a period of two years. We wish to emphasize that a relatively small number of accounts have created and amplified a significant volume of hateful speech in a limited time period, encompassing numerous themes and utilizing various strategies to avoid detection. Consequently, the widespread adoption of ChatGPT in the Indian context may well result in an explosion of such content with grave implications for India's religious minorities, including threats, psychological harm, and physical violence. The large-scale dissemination of hateful

content also carries serious risks of damaging social relations between groups, undermining the principles of constitutional secularism, and attenuating both democratic institutions and the spirit of democracy in Indian society and culture. The proliferation of hateful AI-generated content threatens to further colonize the Indian information sphere, which is already marked by rampant misinformation, anti-minority bias, and a severe crisis of credibility.

The structure of the report unfolds as follows. First, we discuss the global and national context in which AI-generated hate content circulates. Next, we detail our research methodology, including processes of data collection and verification. Following a summary of key findings, the core sections present an analysis of dominant narrative trends such as conspiratorial, exclusionary, sexualized, and aestheticized content. This is followed by a semiotic analysis of the images and an examination of the Hindu far-right media's role in amplifying the images. The report concludes with a reflection on the broader implications of AI-enabled hate production in India and a set of recommendations for stakeholders.

1.1 HATEFUL USE OF AI-GENERATED IMAGES

AI-generated images are pictures produced entirely by software rather than cameras. A generative model, typically a diffusion system like Stable Diffusion or DALL·E, takes a text prompt, starts with random noise, and iteratively "denoises" it by applying patterns it has learned from billions of training photos until a brand-new image that has never existed before appears, often in just a few seconds, on-screen on consumer hardware like a laptop or cellphone. Google's psychedelic "DeepDream" experiments[5] briefly captivated the internet in 2015, but AI-generated images only became a mass phenomenon after mid-2022. Midjourney opened its Discord beta on July 12, 2022. Stable Diffusion's open-source model was released on August 22, 2022, and amassed roughly 10 million users within two months. OpenAI lifted the DALL·E 2 wait-list on September 28, 2022, at which point 1.5 million people[6] were already creating more than two million pictures daily.

As a country with an estimated 900 million internet users[7], Indians have played an important role in engaging with and shaping the landscape of AI-generated imagery. A survey by LocalCircles[8] found that nine percent of Indian AI-platform users, numbering about 22 million, employ tools "for creating or enhancing pictures." While much of this use is innocuous and routine, everyday activity, a growing climate of hate, intolerance, and violence in the country[9] means that AI image generation is being used to target religious minorities, particularly Muslims, Christians, and other marginalized Indian groups like caste-oppressed communities.

This report examines the state of AI-generated hateful visual content in India. We use the term "AI-generated hateful content" to cover all images and videos produced with generative AI tools that (1) target a protected group with negative stereotyping, dehumanization, or incitement to harm, and (2) are disseminated with the reasonably foreseeable effect of amplifying hostility against a protected group. Protected groups in this context refer to communities historically vulnerable to structural discrimination and violence, such as those

marginalized by religion, caste, gender, or ethnicity. In the Indian context, these groups include religious minorities, such as Muslims, Christians, and Sikhs, as well as caste-oppressed groups such as Dalits. Although we note international parallels in the patterns analyzed in the report, the empirical analysis undertaken in the report centers on posts related to the Indian context.

The rapid advancement of Artificial Intelligence (AI) text-to-image technology has introduced a new set of challenges for the global landscape of disinformation and extremism. AI text-to-image generation has made creating photorealistic disinformation cheap, fast, and deployable against specific targets. The Royal United Services Institute's June 2025 report, Online Hate Speech and Discrimination in the Age of AI[10], shows that generative AI has become a force multiplier for online hate, allowing extremist actors to produce synthetic visuals that align neatly with existing prejudices.

Far-right parties and actors across Western Europe are increasingly using[11] text-to-image technology to fuel fear-based narratives. After the 2024 Southport stabbings[12] in the U.K., which involved a teenager killing three young girls, AI-generated images were used to stoke unrest and spread Islamophobic and anti-immigrant hate online based on false claims about the attacker's identity. In April 2025, Italy's opposition parties filed a complaint[13] against Deputy Prime Minister Matteo Salvini's League party for distributing AI-generated images depicting men of color assaulting women or police, based on spurious connections with immigration crime reports. The vilification of people of color and immigrants is a persistent trope in the increasing use of AI-generated imagery by the global far-right. A European Digital Media Observatory (EDMO) investigation[14] found that across Europe, far-right parties and influencers from Germany's Alternative für Deutschland (AfD) to Poland's Sovereign Poland to France's Reconquest, Italy's League and Ireland's The Irish People have used AI-generated images and videos that juxtapose "ideal" and idyllic white communities with ominous-looking and threatening dark-skinned migrants to stoke xenophobic fears.

A key dynamic underpinning this development is the rise of a phenomenon called slopaganda[15], that is, the deliberate use of cheap, abundant, and low veracity synthetic content to seed hate across digital spaces. Such AI slop proliferates in significant measure because generative tools make the creation and sharing of such content easy, allowing its producers to flood platforms and crowd out other content. These tools enable producers of slopaganda to game the algorithmic principles[16] that are key to the major platforms. Far-right organizations across the globe also use these tactics to increase the visibility of extremist content. They are adept at using popular filters and trends[17], often utilizing a commonly recognizable aesthetic language to boost the popularity of such content.

Along with xenophobic propaganda, AI-generated imagery has been harnessed to fuel misogynistic hate. In Australia, eSafety Commissioner Julie Inman Grant flagged[18] a sharp rise in AI-generated sexual deepfakes circulating among students. She urged schools[19] to mandatorily report such incidents under new state laws, with the threat of potential fines up

to $49.5 million per breach. A survey of over 16,000 respondents[20] across ten countries found that 2.2 percent of the survey participants had been victimized by deepfake pornography and 1.8 percent admitted to circulating such content, revealing that existing non-consensual imagery laws are insufficient deterrents to the problem. Indeed, the combination of different kinds of prejudice in hateful speech is a common feature of much hateful content generated through AI tools.

1.2 ANTI-MUSLIM HATE IN INDIA

Global trends related to the misuse of AI-generated imagery are alarming enough, but India presents a particularly worrisome case of the phenomenon. Over the past decade, anti-Muslim sentiment in India has strengthened and is increasingly visible across mainstream political discourse, media narratives, and digital platforms. Anti-Muslim sentiments have manifested themselves in various forms, including targeted sectarian violence, mob lynchings, inflammatory hate speeches, forced evictions of Muslims, destruction of Muslim properties and places of worship, economic exclusion and marginalization, and the normalization of conspiratorial rhetoric that depicts Muslims as outsiders, security threats, or a demographic danger to the Indian nation.

India possesses one of the world's largest and most active digital ecosystems. This vast digital infrastructure enables content to be produced, shared, and consumed at unprecedented speed and volume across multiple platforms and in multiple languages. A deeply troubling aspect of the Indian online ecosystem is the use of digital tools and platforms to attack historically marginalized communities, such as Muslims, Christians, Dalits, and indigenous groups, as well as journalists, academics, and civil society actors. Though reflective of a global trend, the abuse of digital tools has taken on radically dangerous proportions in India.

A report[21] by the Center for the Study of Organized Hate (CSOH), published in February 2025, showed that, of the 1,165 hate speech events CSOH researchers had recorded in 2024, as many as 995 were first shared or live-streamed on social media platforms. Another CSOH report exposed how cow vigilante groups[22] were using Instagram to organize, fundraise, incite, and glorify violence against Muslims. YouTube has similarly been used to host[23] 'Hindutva pop' music that glorifies extremism and targets Muslims. Social media platforms have played a significant role in the propagation and normalization of hate speech and disinformation against Muslims. In 2021, a Meta employee and whistle-blower Frances Haugen revealed compelling evidence that the organization was aware of the fact that users in India were being flooded[24] with hateful propaganda and disinformation against Muslims, which was instrumental in fueling hate and violence against the community.

Given this situation, AI-generated hate content and disinformation have found fertile ground in the country, routinely targeting India's approximately 200 million[25] Muslim population. This report focuses on one specific facet of this phenomenon: the deployment of generative AI to produce and spread anti-Muslim visual hate content online.

2. METHODOLOGY AND DATA COLLECTION

This study employed a multi-step methodology to gather data that blended purposive qualitative sampling with search and verification technologies to confirm the authenticity of data and track its spread.

In step 1, employing a method of purposive sampling, we identified 297 accounts across various social media platforms, including X, Instagram, and Facebook, for further analysis. These 297 accounts had a well-documented and readily available history of posting harmful content that expressly targeted India's religious minorities, especially Muslims and Christians. These accounts were selected from CSOH's internal dashboard on hate, which tracks accounts that regularly post harmful content.

We used two basic criteria to compile the list of 297 accounts.

1. We utilized the definition of hate speech proposed by the UN Strategy and Plan of Action on Hate Speech[26] to select accounts whose content met the relevant criteria. Per the UN document, hate speech refers to "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor." The 297 accounts that most clearly and persistently met this definition were chosen by a process of selection and vetting by a CSOH research team.
2. Content from these accounts incorporated AI-generated synthetic images to convey harmful sentiments, primarily through the image itself or in combination with written text and captions. Given some accounts' longevity and the relative recency of adopting AI-generation tools, harmful content churned out by these accounts includes, but is not limited to, material that uses AI-generated synthetic images. These accounts have a longer history and a broader range of producing hateful content that precedes and exceeds the use of AI-generated synthetic images. The accounts also continue to produce content that does not include AI-generated synthetic images.

Step 2 of the methodological process involved manually examining the 297 accounts for posts produced between May 2023 and May 2025 that included AI-generated harmful content aimed at Muslims. We focused on posts in Hindi and English across X (formerly Twitter), Facebook, and Instagram. We found most posts using AI content to be concentrated from January 2024 to April 2025.

This process included scrolling through timelines, reviewing media attachments, reading accompanying captions, hashtags, and comment threads, and identifying posts that explicitly or implicitly targeted Muslims using AI-generated imagery. We observed that these far-right accounts relied on open-source diffusion models or other uncensored AI applications, where filters could be bypassed. We noted that their use to spread extremist content picked up after

2023, mainly due to a surge in the popularity of such models in India around that time. Many of these images were subsequently reposted and circulated by other accounts, which streamlined their visibility across Hindu far-right and media networks and reinforced the overall rise in volume of hateful content. Step 2 yielded 1,326 publicly available AI-generated images retrieved from the 297 public accounts.

Step 3 involved a cycle of manual coding across the 1,326 images to first identify broad categories and subcategories of AI-generated anti–Muslim hateful content across the archive. At least two members of the CSOH research team undertook the coding process, identifying themes and categories and reconciling findings to ensure inter-coder reliability. Based on the results of the coding, we classified our findings into four main categories: conspiratorial narratives, exclusionary and dehumanizing rhetoric, the aestheticization of violence and the sexualization of Muslim women.

Step 4 of the process centered on tracking the spread of narratives in the archive of hateful visual content. To do this, we searched for identical captions on X, Facebook, and Instagram, ran Google Reverse Image Search, and manually tracked reposts across platforms. A recurring pattern emerged: when a visually compelling AI image was posted, many accounts recycled the identical content, often copying the original caption. During this phase of data gathering, we observed frequent use of AI-generated imagery by Indian far-right news outlets such as Sudarshan News, Swarajya, Panchjanya, Kreately.in, Organiser Weekly, OpIndia (English and Hindi), Aapka Bharat, Satyaagrah, and Zee News (English and Hindi). We completed the analysis by June 15, 2025. Any changes that may have been made to the original sources that constituted our dataset were not recorded after this point.

Step 5 of the methodological process focused on verifying the images to ensure data credibility. Sight Engine[27], a computer-vision service that flags synthetic visuals and attributes them to their probable text-to-image generators, was used for image verification. Each image in the dataset (or a screenshot of the same) was scanned, and those receiving high-confidence AI scores (e.g., "99 % likely to be AI-generated") were annotated accordingly. A 2024 benchmarking study[28] ranked Sight Engine above competing models for accuracy, precision, and tool attribution, so we selected it as our verifier for this study.

2.1 SCOPE AND LIMITATIONS OF THE STUDY

While this report does not purport to be exhaustive, it offers critical insight into a rapidly evolving phenomenon and presents an early intervention in understanding the use of generative AI for the production and circulation of anti-Muslim hate in India. It also bears value for understanding the use of the technology for the spread of anti-minority hate within and across diverse sociopolitical contexts. The report combines qualitative depth, visual semiotic analysis, and platform-level documentation. It offers a robust foundation for further investigation into the intersection of AI technologies and targeted hate.

The dataset is limited to purposively sampled and publicly available posts in English and Hindi on X, Facebook, and Instagram. Content shared in private groups, deleted prior to collection, or posted in various vernacular languages is not included in the sample. While we used detection tools such as Sight Engine for capturing AI-generated images, some images, particularly those that may have been subtly manipulated or edited to bypass filters, may have been missed. Engagement metrics were recorded at the time of collection and may not reflect the long-term reach of posts.

While the dataset captures a narrow slice of the broader information ecosystem and, accordingly, represents a small sample of the total volume of AI-generated content circulating online, it provides a snapshot of the thematic, semiotic, aesthetic, and strategic dimensions of the use of AI-generated visual content for the expression of hateful sentiments. The volume, consistency, and repetitiveness of the content identified by the research team are indicative of a much larger and expanding apparatus of AI-generated hate. Even in the brief duration between the time of data collection and the publication of this report, the CSOH research team observed several posts gaining steadily increasing engagement.

The patterns documented in this report thus reflect broader trends in the weaponization of AI for targeted hate and propaganda, which merit urgent and sustained investigation as the technologies themselves, as well as the strategies of how they are deployed, continue to evolve. This report helps lay the groundwork for future studies, policy responses, and platform accountability efforts aimed at mitigating AI-enabled harms. It also offers civil society actors, researchers, and regulators a starting point to better understand and respond to the emerging threats posed by generative AI in the context of religious hate and disinformation.

3. KEY FINDINGS

- Generative Artificial Intelligence (GAI) tools are starting to play a central role in the production of anti-Muslim hate in India, allowing widely accessible technology to generate and amplify Islamophobic narratives at an unprecedented scale.
- We identified 1,326 AI-generated hateful posts targeting Muslims between May 2023 and May 2025 from 297 accounts across X, Instagram, and Facebook that were specifically selected for this study. Activity was minimal in 2023 and early 2024 but rose sharply mid-2024 onward, reflecting the growing popularity of and easier access to AI tools that accelerated the rapid creation of hateful content.
- Instagram drove the highest engagement (a sum of likes, comments, and shares) with 1.8M interactions across 462 posts, compared to 772.4K on X (509 posts) and 143.2K on Facebook (355 posts). Instagram emerged as the most effective amplifier of AI-generated hate despite hosting fewer posts than the other platforms.
- The AI-generated hateful content clustered around four dominant themes: the sexualization of Muslim women, exclusionary and dehumanizing rhetoric, conspiratorial narratives, and the aestheticization of violence.
- The category of sexualized depictions of Muslim women received the highest engagement (6.7M interactions), revealing the gendered character of much Islamophobic propaganda, which fuses misogyny with anti-Muslim hate.
- Conspiracy theories such as 'Love Jihad,' 'Population Jihad,' and 'Rail Jihad' were widely reinforced through AI-generated imagery, framing Muslims as a perpetual threat to Hindu society and national security.
- AI-generated images depicted Muslims as snakes wearing skullcaps, a dehumanizing metaphor that framed them as deceptive, dangerous, and deserving of elimination.
- Stylized and animated AI aesthetics (including Studio Ghibli–style imagery) made violent, hateful content appear palatable, even humorous, broadening its reach among younger audiences.
- Hindu nationalist media outlets, notably OpIndia, Sudarshan News, and Panchjanya, played a central role in producing and amplifying synthetic hate, embedding AI-generated Islamophobia into mainstream discourse.
- Across X, Facebook, and Instagram, 187 posts were reported for violating community guidelines. Of these, only one was removed on X, demonstrating the platforms' consistent failure to enforce their own policies.

4. DATA ANALYSIS

The report analyzed 1,326 posts from 297 accounts, including 146 accounts on X (formerly Twitter), 92 accounts on Instagram, and 59 accounts on Facebook, related to AI-generated anti-Muslim hate imagery and narratives. Of these accounts, 86 were verified. The 1,326 posts were selected because they (1) used AI-generated visuals and (2) contained explicit hateful references to Muslim identity, based on the UN definition employed in the report.

These posts received a total engagement of 27.3 million across X, Instagram, and Facebook. Of the 1,326 posts analyzed, 509 were from X, 462 from Instagram, and 355 from Facebook, accounting for 24.9 million engagements on X, 2.32 million on Instagram, and 152,100 on Facebook.

Engagement was calculated by summing up the available interactions on each platform.

- On X, this included views, likes, reposts, and comments;
- On Facebook, this included likes, shares, and comments;
- On Instagram, this included likes and comments; shares were included where available.

Total recorded engagement across the verified corpus was 27.3 million, comprising platform-reported views, likes, shares, and comments.

4.1 PLATFORM BREAKDOWN

- The posts in our dataset drew 24.5 million views, 2.2 million likes, 70.1k comments, and 501.1k shares across X, Instagram, and Facebook.
- These posts drew 24.9 million total engagements on X, 2.32 million on Instagram, and 152.1k on Facebook, with X accounting for the overwhelming share.
- Across platforms, X received 24.1 million views, 564.4k likes, 29.2k comments, and 179.8k shares; Instagram saw 1.53 million likes, 24.8k comments, and 313.1k shares; while Facebook had 119k likes, 16k comments, and 8.2k shares.

TABLE 1: AGGREGATE ENGAGEMENT METRICS ACROSS ALL PLATFORMS

Total Views	24.5 M
Total Likes	2.2 M
Total Comments	70.1K
Total Shares	501.1K
Total Engagement	27.3M

TABLE 2: PLATFORM-WISE ENGAGEMENT BREAKDOWN OF AI-GENERATED POSTS

Platform	Likes	Comments	Shares	Total Engagement (Except Views)	Total Number of posts
X	564.4K	29.2K	179.8K	772.4K	509
Instagram	1.5M	24.8K	313.1K	1.8M	462
Facebook	119.0K	16.0K	8.2K	143.2K	355

Note: Views were excluded as they are not consistently available across all platforms.

4.2 MAIN THEMES

Each post was manually reviewed and assigned a primary thematic tag based on its visual or textual content. Five main trends emerged, with some overlap, that captured the majority of posts analyzed:

1. **Conspiratorial Islamophobic Narratives:** These narratives broadly suggest that Muslims as a community are engaged in conspiracies to undermine the national integrity or security of India. In posts, conspiracy theories about Muslims were frequently yoked to claims about Muslim sexual mores or the natural disposition of the community towards violence.

2. **Exclusionary and Dehumanizing Rhetoric:** This category consists of images that suggested or encouraged violence against Muslims. Images emphasize the dehumanization of Muslims through a number of tropes, including the use of animal imagery to describe the community, the ascription of Muslim hatred directed at non-Muslims, and the attribution of a desire on the part of Muslims to sabotage Indian property and undermine national interests.

3. **Sexualization of Muslim Women:** This category of images represented Muslim women in a sexualized manner, dehumanizing and rendering them as legitimate objects of sexual violence.

4. **Aestheticization of Violent Imagery:** This category of images used AI to link historical, conventional, or everyday images of Muslims with incidents of sectarian anti-Muslim violence, often with the effect of normalizing such violence.

5. **Other:** A number of posts that targeted Muslims using AI-generated imagery without any tentative thematic connections to the above-described categories. These images were classified as 'other.'

FIGURE 1: DISTRIBUTION OF POSTS BY MAIN THEME

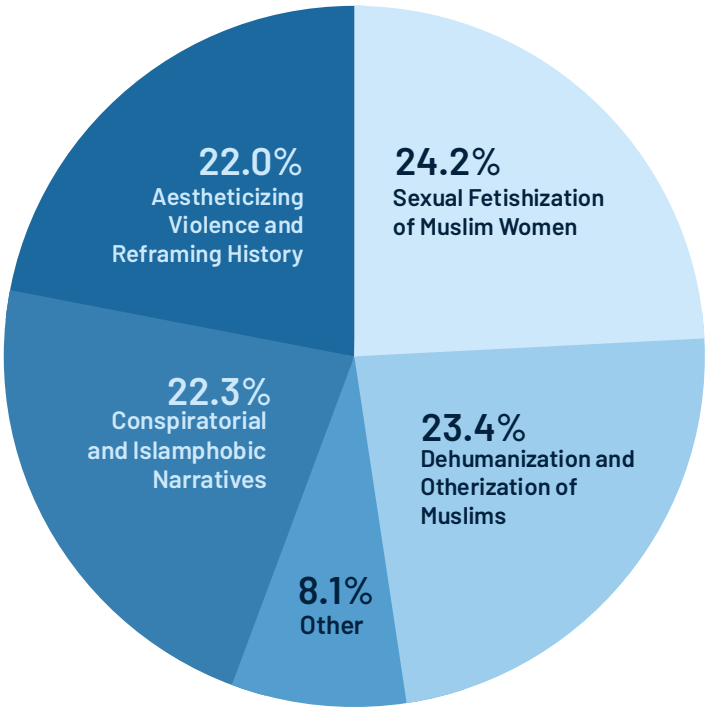


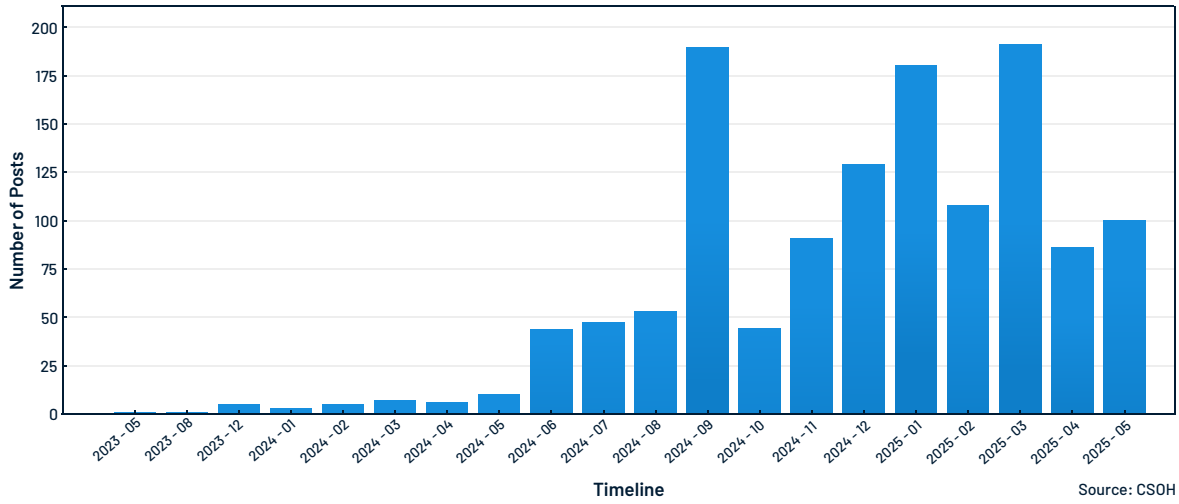
TABLE 3: BREAKDOWN OF THEMES WITH ASSOCIATED ENGAGEMENT

Narrative	Posts	Views	Likes	Comment	Share	Total Engagement
Aestheticizing Violence and Reframing History	134	5.15M	741K	12.8K	93.5K	6.0M
Conspiratorial and Islamophobic Narratives	457	5.1M	695.2K	15.4K	236.4K	6.0M
Dehumanization and Exclusion (merged category)	291	5.79M	462.28K	18.84K	130.94K	6.4M
Sexual Fetishization of Muslim Women	317	6.58M	89.82K	11.54K	9.86K	6.7M
Other	132	1.9M	231.4K	11.6K	30.3K	2.2M

While the timeline of the study spans two years, most AI-generated harmful posts were concentrated within a shorter period, particularly from mid-2024 onward. Activity was minimal through 2023 and the first half of 2024, but began rising sharply in June 2024, with visible spikes in September 2024 and March 2025. The September spike coincided with the circulation of content around the 'Rail Jihad' conspiracy theory, while the March surge reflected the popularity of the Ghibli Art trend in AI-generated hateful imagery.

This pattern shows that the use of AI imagery to promote Islamophobic narratives only gained traction in 2024, marking a turning point in both volume and sophistication of content within the Hindu nationalist ecosystem. A key driver of this shift was the growing popularity and wider availability of advanced text-to-image tools, which made producing visuals easier. OpenAI's integration of DALL·E 3 into ChatGPT in September 2023 likely accelerated the adoption of such tools, lowering barriers to use and fueling their incorporation into propaganda at scale.

FIGURE 2: NUMBER OF POST OVERTIME (2023 – 2025)



5. NARRATIVE, THEMATIC, AND STYLISTIC ANALYSIS

As briefly described above, the dataset includes distinct and recurring narratives, themes, and aesthetic idioms that show how AI-generated images are being strategically deployed to target Muslims in India. Such usage is reflective of an emerging far-right tactic that weaponizes AI to shape public perception, reinforce existing prejudices, and lay the ground for real-world harm against Muslims. In this section, we detail these narratives, themes, and aesthetic and stylistic idioms.

5.1 CONSPIRATORIAL ISLAMOPHOBIC NARRATIVES

AI-generated images are being systematically used to repackage well-known Islamophobic tropes into new content as part of various conspiratorial narratives. Anti-Muslim hate in India draws significantly on the essentialist idea of Muslim criminality, which is perpetuated through depictions of Muslims as violent, predatory, and immoral. AI-generated images reinforce this notion of Muslim criminality by combining propagandist conspiracies with visual prompts. The use of images in content is widely understood to help generate greater engagement[29] on social media. AI-generated visual narratives draw from a cluster of conspiracy theories that present Muslims as threats to Hindu women, Hindu families, and the Indian nation itself.

With 457 posts, this category received 5.1M views, 695.2K likes, 15.4K comments, and 236.4K shares, leading to a total engagement of 6 million. We identify three sub-trends within this category: Muslims as inherently violent; Muslims as moral and sexual degenerates; and jihad-linked conspiratorial narratives.

5.1.1 THE PORTRAYAL OF MUSLIMS AS VIOLENT

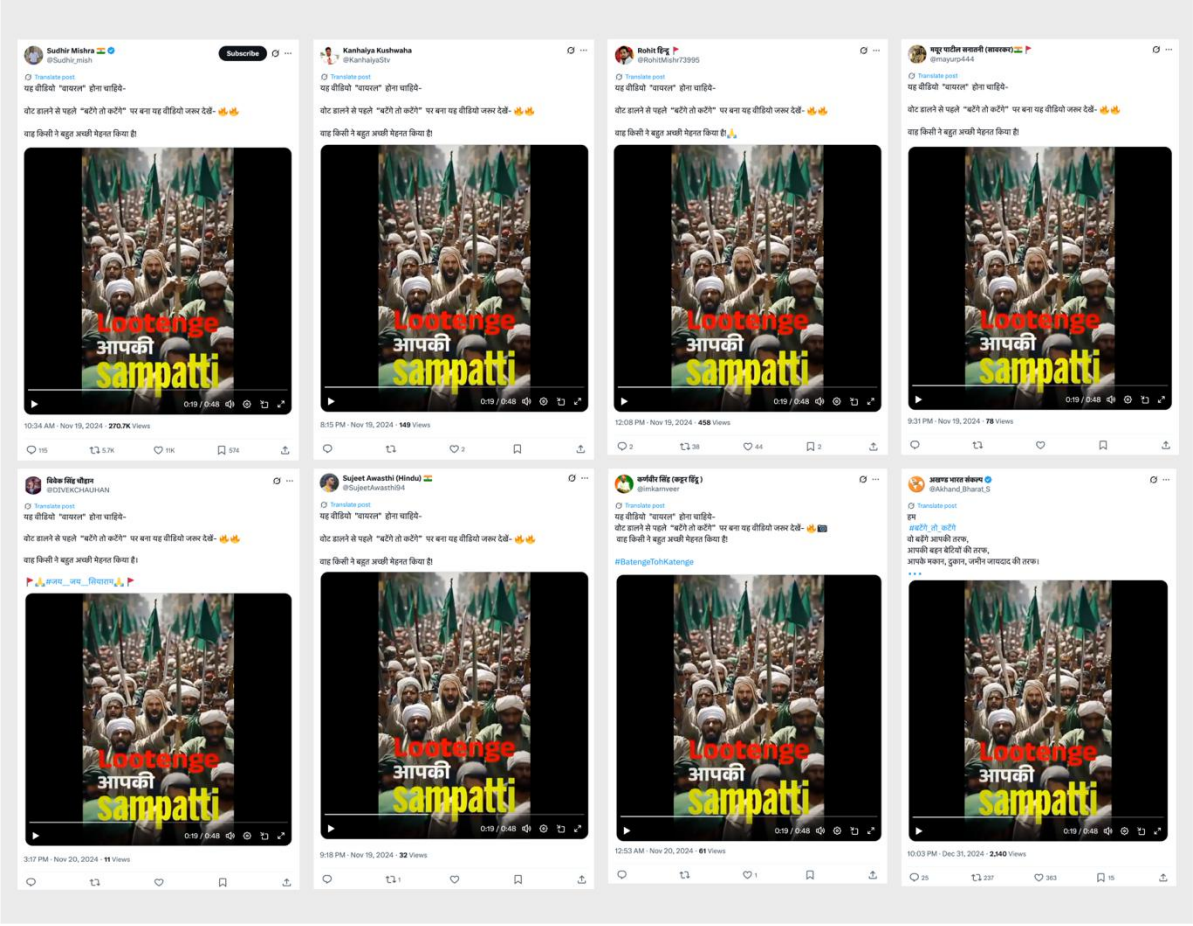
AI-generated images depicting Muslims as an inherently violent community are a recurring tool in digital propaganda. These images depict Muslim men as aggressors, angry, armed, and engaged in various acts of destruction. The images feed into a broader narrative of Muslims as a permanent threat to the safety of Hindus in India.

Many of these images depict scenes of riots, arson, or assaults, with Muslim figures presented as the violent perpetrators of such acts. The images are usually accompanied by text or captions that amplify or bolster the narrative. We find this pattern in posts regarding the alleged killing[30] of 17-year-old Sudipta Pandit by Sheikh Farhad Mondal, a mango orchard owner, in Naihati, West Bengal in May 2025. The killing resulted from a conflict in which Pandit was accused of stealing mangoes by Mondol. Hindu nationalist outlets such as OpIndia[31] and Hindu Post[32] were quick to allege that Sudipta was killed because he was Hindu, despite no evidence of sectarian motivation behind the act. One AI-generated image showed the accused wearing a skull cap, typically worn by Muslim men, in order to imply anti-Hindu motivation for the murder. The choice to incorporate Muslim-coded clothing in the image may be

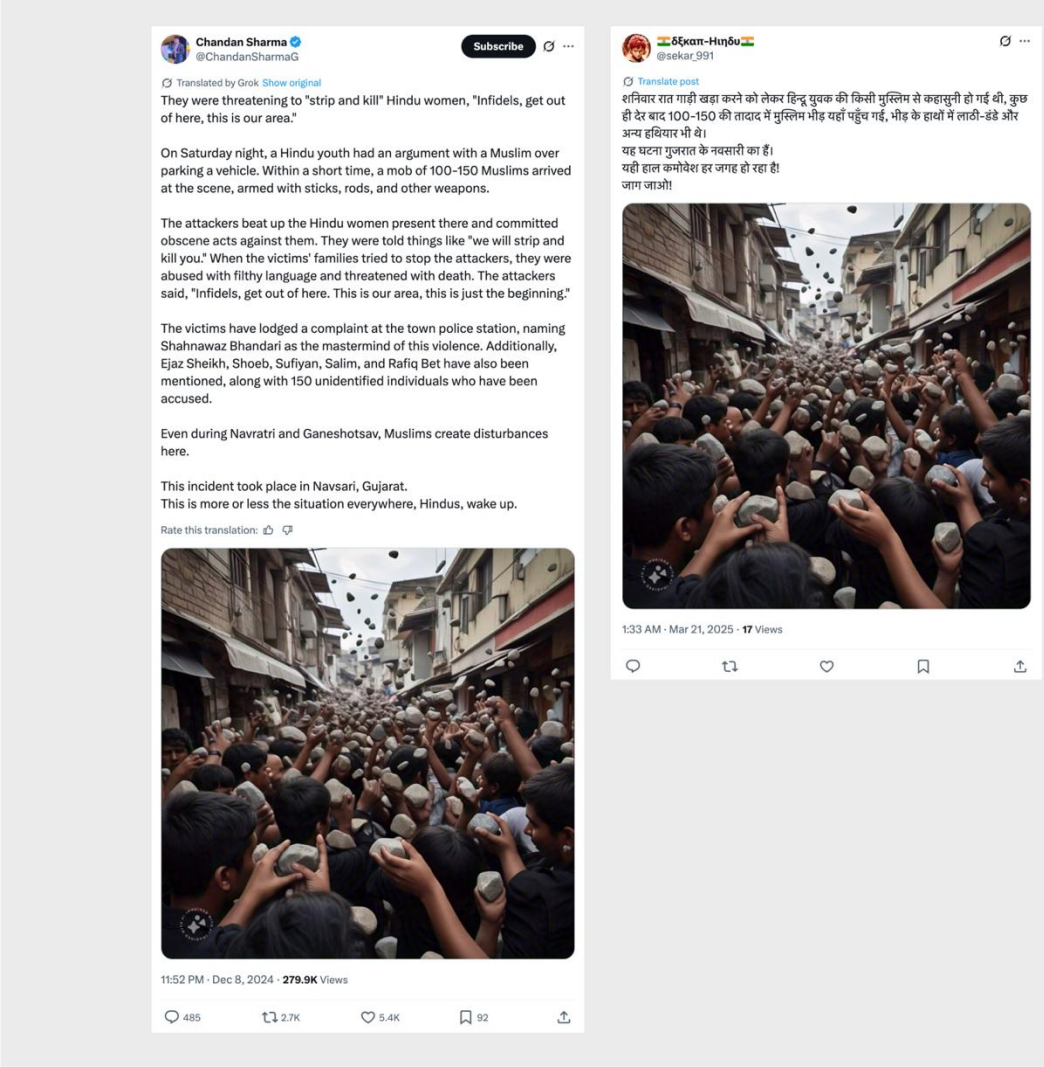
inferred as deliberate, as none of the news reports stated what Sheikh Farhad was wearing during the incident.

The use of AI-generated imagery allows Hindu nationalist actors to present hateful narratives without the possibility of either being fact-checked or flagged through a community note. We find such content syndicated and used across multiple platforms by different users. One video, included below, that is composed entirely of AI-generated images, and was repeatedly shared across Instagram, X, and Facebook, contrasts depictions of aggressive Muslim men with portrayals of Hindus as vulnerable.

The Hindu identity is figuratively depicted in the form of the elderly, women, and young men with folded hands. The video is accompanied by the slogan “Batenge to katenge” (“If divided, we will be destroyed”), originally popularized by Yogi Adityanath, the Hindu nationalist Chief Minister of the Indian state of Uttar Pradesh[33], during general and state elections in 2024. The video relies entirely on AI-generated images of Hindu victimhood and Muslim aggression, while the viewer also listens to hateful speech that pairs with the visual narrative.



In a politically fraught climate, such as in present-day India, common incidents of protests or localized conflicts in neighborhoods can quite easily be reframed through the lens of ethnic or sectarian strife by bad-faith actors. AI-generated images can be conveniently mobilized along these lines. A parking dispute between a Hindu man and a Muslim man in Navsari, Gujarat, on December 8, 2024, was spun as sectarian conflict on social media in this very manner. A post[34] on X claimed that, following the disagreement, a crowd of 100 to 150 Muslims armed with sticks and other weapons had descended on the area, assaulted Hindu women, and made obscene gestures. The Muslims had allegedly threatened to "strip and kill" the women, told "kafirs," or infidel non-Muslims, to leave because it was their area, and had warned Hindus that such actions were "just the beginning" of what they might do to the latter. The post also claimed that police complaints had named several Muslim men as masterminds behind the threats and violence and that similar "Muslim trouble" occurred during Hindu festivals such as Navratri and Ganeshotsav. The accompanying image in the post shows a crowd of men with stones in their hands, creating the impression of an organized mob of Muslims baying for Hindu blood.



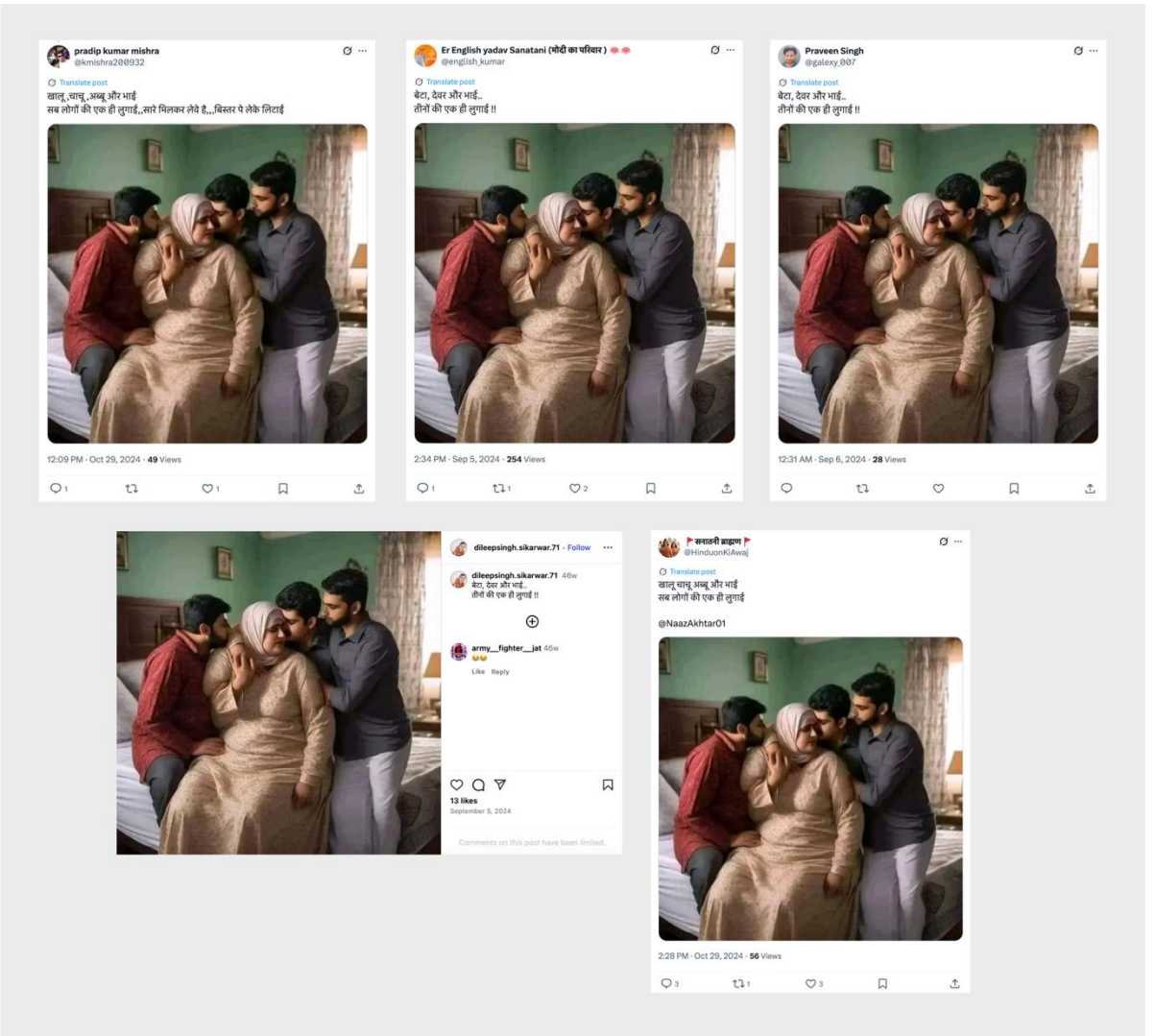
However, according to the police, the incident had no sectarian angle. It was strictly a dispute about parking, and there were no religious slurs or threats made by any Muslim. Fact-checking outlets[35] also debunked the possibility of any angle of religious conflict in relation to the incident and confirmed the absence of any targeted violence against Hindus by Muslims.

5.1.2 SEXUAL AND MORAL DEGENERACY OF MUSLIMS

There is significant diversity of cultural and familial practices within and across religious communities in India. Muslims and Hindus are also governed by separate personal laws in the realm of marriage practices and inheritance. A professed objective of the Hindu far-right is the imposition of a Uniform Civil Code[36] in India, which would standardize personal laws across religious groups. While there is a case for a uniform civil code, such as, for instance, on grounds of gender equity, it is essential to note that the Hindu far-right continues to use the issue as a political plank and as a pretext for denigrating and attacking Muslims. Many Hindus and Christians also oppose a uniform civil code.

In order to drum up support for a uniform civil code, Hindu nationalists have long accused Muslims of perverse familial practices, often exaggerating their differences[37] from normative Hindu practices. AI-generated images echo this pattern, depicting Muslim women in burqas surrounded by groups of men, accompanied by captions implying incestuous family relationships. These images vilify Muslims as a community defined by aberrant social and familial customs.

The use of the burqa in such fabricated images serves a dual purpose. One, it underscores the idea of Muslim women as lacking agency, with their existence determined entirely by the oppressive patriarchal environments that they inhabit. And, two, the burqa functions as a symbol of moral deviance. The captions accompanying the images play a key role in conveying these messages. They frequently contain unverified claims about forced marriages, incest, or the suppression of women's rights within Muslim families. In some instances, these particular claims are linked to broader narratives about the nature or character of Islam, such as the idea that Islamic law permits or encourages such practices.



Alongside more explicit accusations of immorality and degeneracy, AI-generated images also employ crude humor and crass visual satire to reinforce degrading stereotypes about Muslim family life. These are often accompanied by captions laced with sexual innuendo.

For instance, Instagram user @shobhit4494 posted an image of a visibly pregnant woman standing beside a man in traditional Muslim attire, who appears confused and says, “Main toh idhar tha hi nahi aur maine kheti bhi nahi ki thi… phir ye kele ki fasal kaise aa gayi?” (I wasn’t even here, nor did I do any farming… then how has this banana crop grown?). The woman responds, “Aapke Abbu, bhai aur padosiyon ne bahut mehnat ki… aur kuch upar wale ki bhi den hai.” (Your father, brother, and neighbours worked very hard… and to an extent it is God’s gift too.)

The image and caption connote incest, infidelity, sexual profligacy, and doubts about the paternity of the child through metaphors of farming, agriculture, and harvesting fruit. While not explicitly mentioning sexual activity, the visual and verbal cues clearly suggest that Muslims are sexually promiscuous and lack any ethical or moral code in their private lives.

Another Instagram post by the same user uses the metaphor of cricket, a sport deeply embedded in Indian popular culture, as a euphemism for sexual acts within the household. One shows a burqa-clad woman holding a cricket ball, asking the male figures around her: “Abbu, aaj batting aap pehle karoge ya bhaijaan?” (Father, will you bat first today or will brother?).

This casually posed question, paired with the phrase "jo bhi kare pitch par aa jao," (whoever will bat first can step up to the pitch), insinuates an incestuous domestic arrangement in which the woman has sexual relations with several male family members. The burqa, once again, functions simultaneously as a marker of Muslim identity and sexual deviance.

These examples reveal how text and visuals collaborate to suggest moral deviance without ever stating so categorically. The coded language serves to protect the posters from direct platform moderation while still communicating messages about deviancy unequivocally to audiences familiar with the cultural subtext. The use of comedy hashtags (#comedy, #prank, #joke, #fun) attempts to sanitize the implications of such content by framing it as harmless humour. But the effect of these images, singly or cumulatively, is the dehumanization of Muslim men as hypersexual predators and Muslim families as incestuous or morally corrupted units.

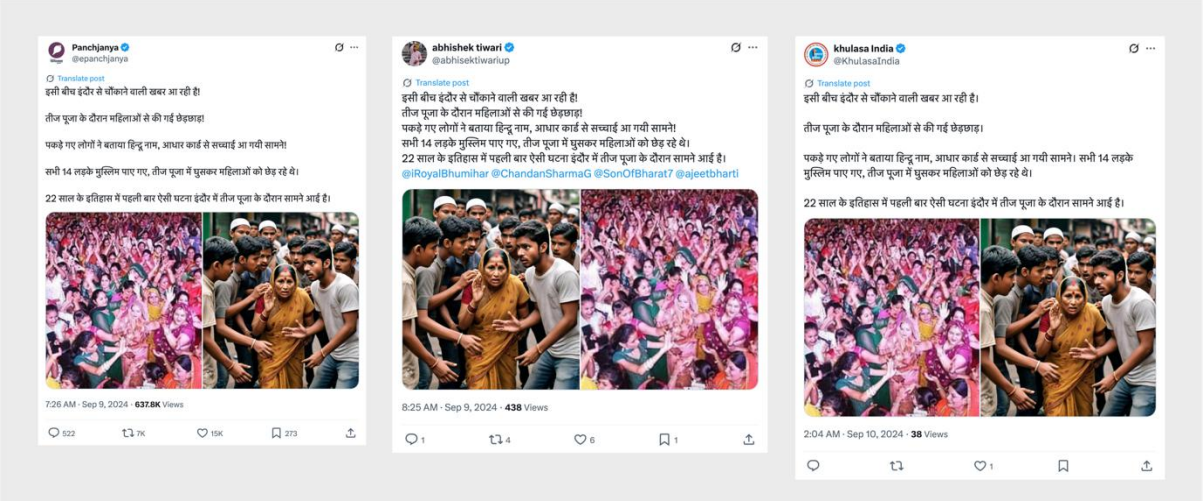
5.1.3 JIHAD-LINKED CONSPIRATORIAL NARRATIVES

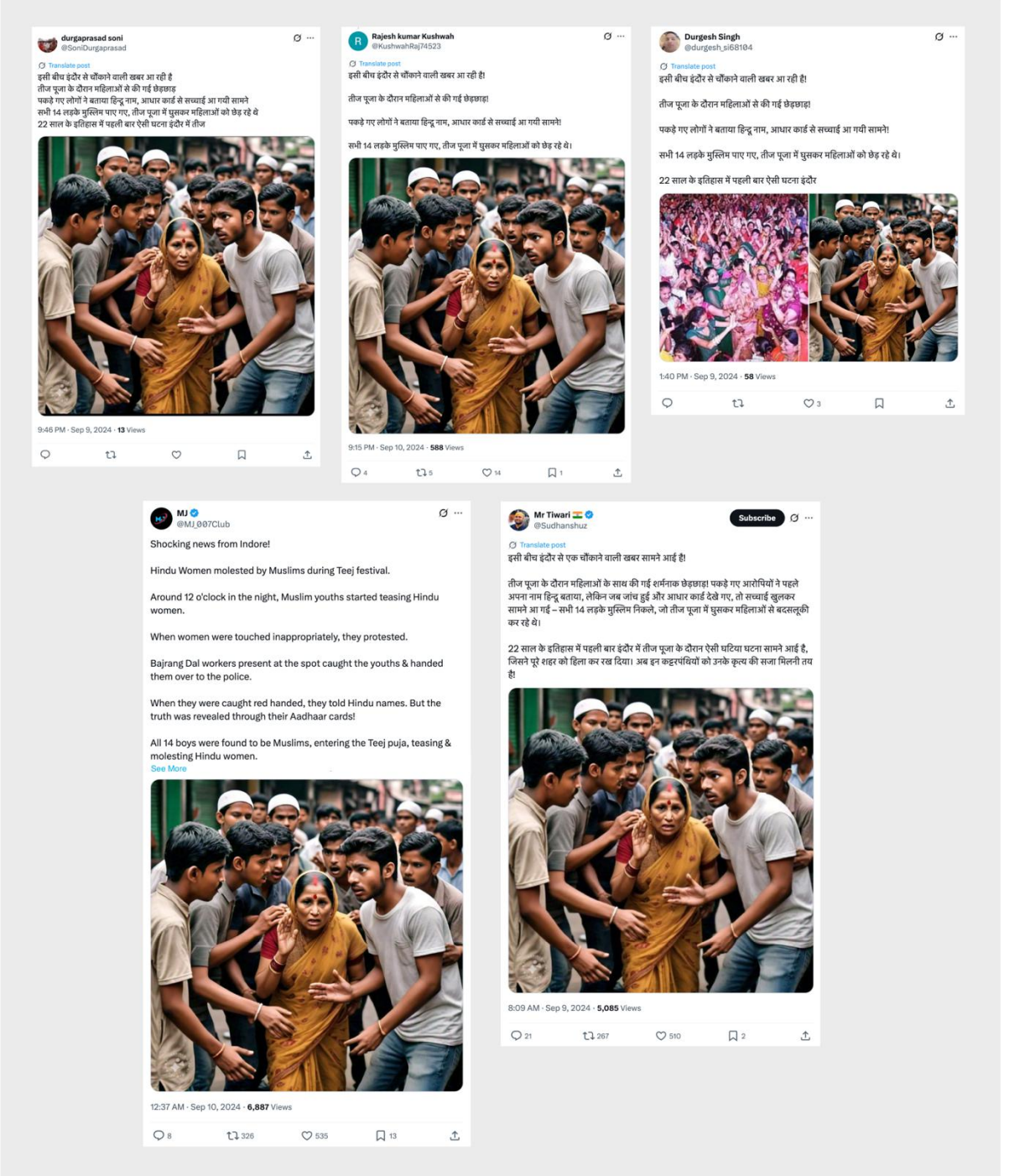
Hindu nationalists have reframed the idea of jihad, linking it to a wider constellation of events and actions, based on the idea that Muslims as a community are committed to engaging in violence against Hindus, specifically Hindu Indians. In tandem, Hindu nationalists denigrate Muslims as a community by alleging that they are guilty of immoral and unethical practices, prone to violence, and engage in subterfuge in order to harm other communities. These claims are often also connected to conspiracy theories about specific forms of jihad, such as 'Love Jihad,' 'Population Jihad,' and 'Rail Jihad.'

LOVE JIHAD

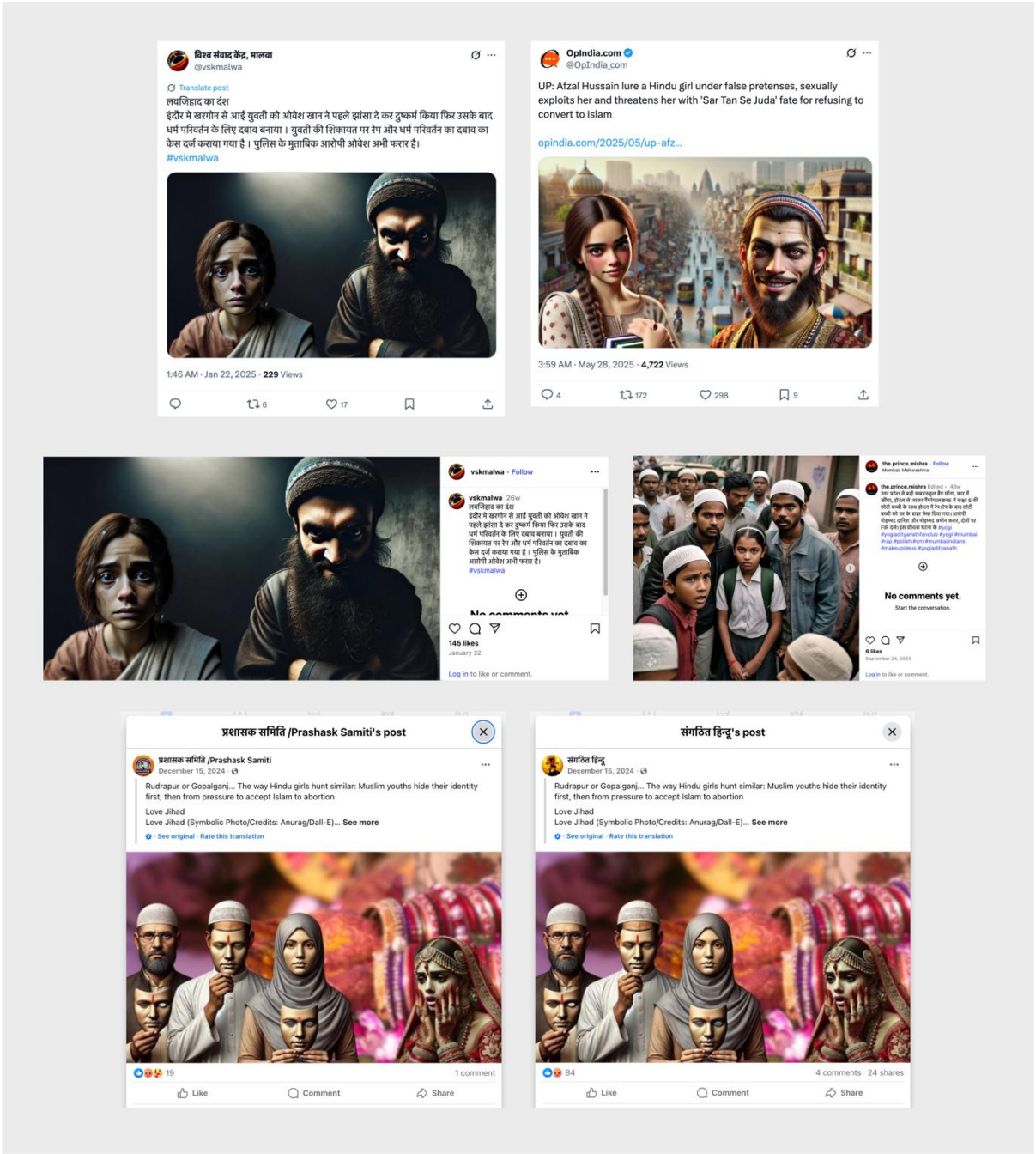
The very existence of Muslims in public spaces and practices is frequently framed as a form of jihad, which itself functions as a synonym for terrorism. Routine and banal everyday practices, including buying property, selling produce, or studying for examinations[38], are presented as acts of terrorism when Muslims engage in them. The most common example of such a rhetorical construction is the "Love Jihad" conspiracy theory, which posits that Muslim men systematically lure Hindu women into romantic relationships with the goal of forcibly converting them to Islam. The Love Jihad conspiracy is invoked to explain a wide range of crimes, including rape and murder. Consensual relationships between Muslims and non-Muslims are also explained as a form of Love Jihad. In the images that accompany posts which center on the idea of Love Jihad, Muslim men are typically shown with menacing expressions while dressed in Muslim-coded attire like skullcaps. These images are juxtaposed with images of frightened or helpless women, usually coded as Hindu, with the contrast reinforcing the trope of Muslim male aggression. AI-generated images on this theme are found in abundance in the archive.

On the night of September 6, 2024, in Indore, Madhya Pradesh, during the Hartalika Teej bhajan sandhya, a religious gathering marking the Hindu festival of Hartalika Teej, a group of eight to ten young men in the crowd allegedly molested several women devotees[39]. Members of the Bajrang Dal, a militant Hindu nationalist group, claimed they had apprehended the men, ascertained that they were Muslims who were using fake Hindu names, and handed them over to the police. The Bajrang Dal did not provide any evidence for the claim that the alleged molesters were Muslims pretending to be Hindus. Online posts related to the event frequently incorporated AI-generated images. For instance, the following image below, shared by several users, shows a woman wearing a tilak, a Hindu religious marker placed on the forehead, who is surrounded by several men in skullcaps, who are clearly meant to be Muslim. The men are shown crowding and intimidating the woman, a metaphor for the assault on Hindu female honor and Hindu identity by Muslims.





AI-generated images that broadly reflect this narrative depict Muslim men, variously, as perpetrators of forced religious conversion, deceptive relationships, and violent threats against Hindu women and minors. These men are portrayed with stereotypical features such as beards, skullcaps, and exaggerated dark facial traits to suggest menace and danger. They are commonly shown engaging in coercion, deception, or violence. Posts sometimes attempt to lend legitimacy to these claims by citing loosely attributed or incomplete legal documents, such as First Information Reports (FIRs).

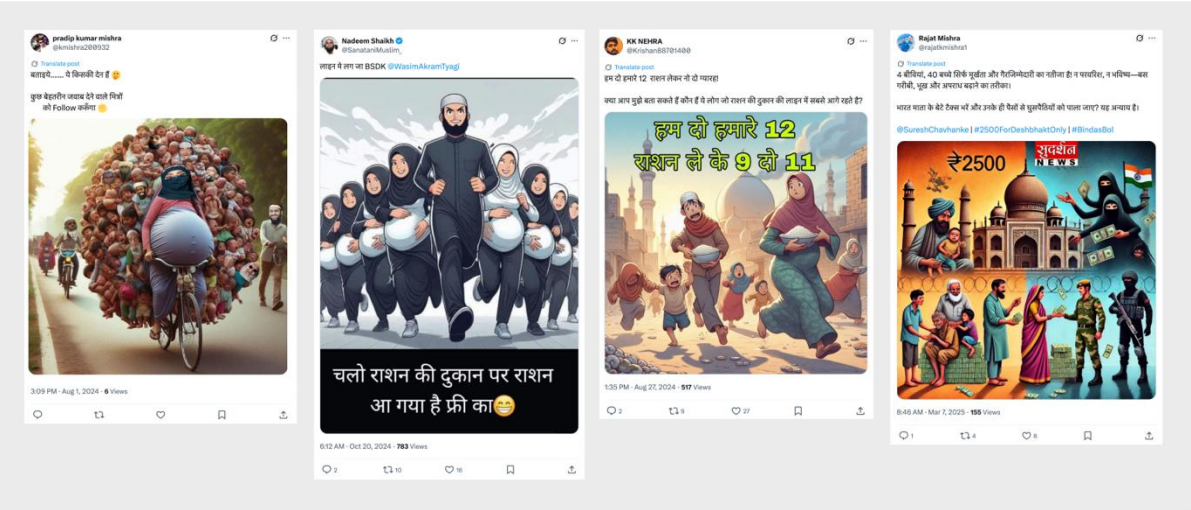


POPULATION JIHAD

AI-generated images are also used to project demographic anxieties about eventual Muslim domination of India through the conspiracy theory of Population Jihad or Demographic Jihad. These particular conspiratorial narratives are anchored in debunked statistical claims that predict that Muslims will eventually outnumber and dominate the Hindu population due to higher fertility rates.[40] This idea is similar to the Great Replacement Theory[41] prevalent in far-right discourse in Western societies, which purports that migrants from the Global South are replacing the ethnic White populations of these countries through higher birth rates and unchecked immigration. In both settings, the theories exploit real and imagined demographic

changes to promote fears of cultural and national decline. They manipulate anxieties about the loss of racial or religious superiority that will upend the dominant social order, framing demographic change as part of a coordinated effort on the part of minorities and immigrants to outnumber supposedly 'original' racial, ethnic, or religious populations.

In the AI-generated images related to this trope in our dataset, Muslim women are depicted as pregnant and/or surrounded by large numbers of children in cramped or impoverished settings, implying unchecked reproduction on the part of Muslims and portraying the community as an economic burden on society and the state. Posts accompanying these images often claim that Muslims exploit India's Public Distribution System (PDS) by using ration cards or other government-issued documents to obtain food at subsidized prices. Such claims construct Muslims not only as a demographic threat but also as a financial liability to the nation.



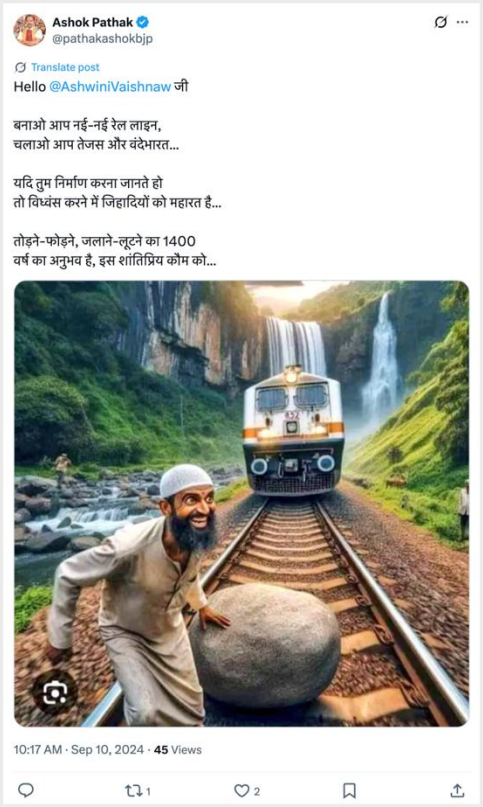
The visual and textual representations of Muslims as a rapidly expanding, economically unproductive, and parasitic population align seamlessly with the narrative of a demographic takeover of India by Muslims.

RAIL JIHAD

Following a series of railway accidents in India, another conspiracy began circulating online: the theory of Rail Jihad. According to proponents of the theory, Muslims were allegedly sabotaging train tracks as part of a coordinated attack on national Indian infrastructure. Though such claims were repeatedly debunked, the conspiracy gained traction with the help of AI-generated visuals, which falsely claimed[42] that rail derailments and accidents had been caused by acts of deliberate sabotage. Many of the spurious claims insist that Muslims[43] are the perpetrators of these acts, with sabotage presented as another form of terrorism.

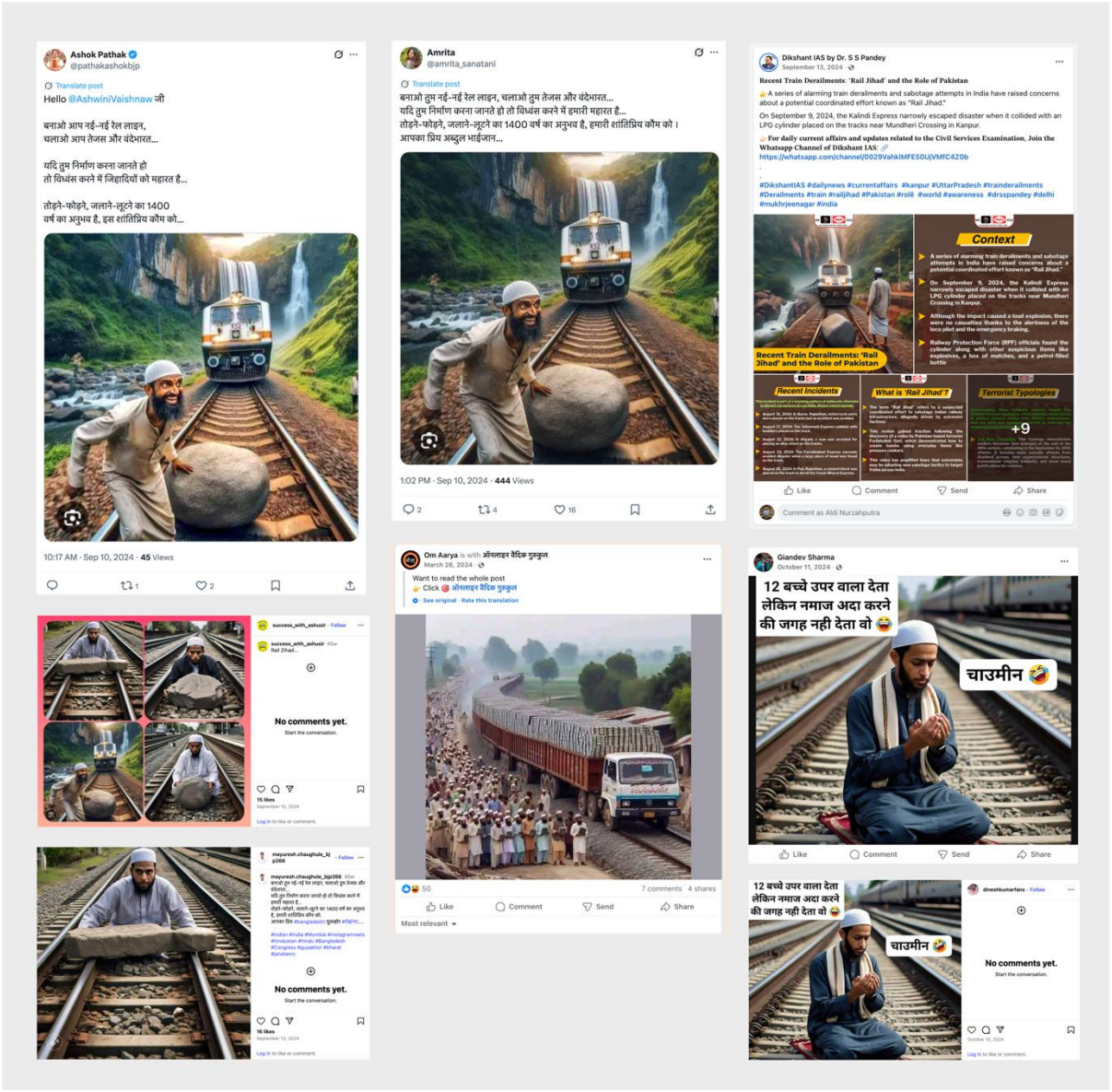
Many videos or pictures[44] that have been shared as proof of Muslims damaging trains[45] or sabotaging railway lines have been debunked as fake. To bypass filters of verification and

fact-checking, users who circulate these bogus reports have started using AI-generated images. Instead of presenting concrete claims about particular incidents that can be checked and countered, the text that accompanies these AI-generated images employs broad generalities, suggesting, for instance, that Muslims are an inherently destructive community who have deliberately been targeting the Indian railway system.



The text in this image paints Muslims as a historically violent and destructive group of people, while sarcastically referring to them as “peace-loving.” The images reference the false claims that Muslims have been placing boulders or other obstacles on railway tracks to derail trains. The term ‘Abdul’ as a generic placeholder for all Muslim men is also a common slur in online Hindu right-wing discourse.

Together, these AI-generated visuals and accompanying narratives transform ordinary railway accidents into evidence of a conspiracy by Muslims, reinforcing stereotypes of Muslims as violent, destructive, and inherently anti-national. The conspiracy of rail jihad has become a recurring theme across social media, with numerous posts recycling the same tropes in different forms.



5.2 EXCLUSIONARY AND DEHUMANIZING RHETORIC

The process of the exclusion and dehumanization of minorities has taken numerous forms through history. Such propaganda typically invites readers and audiences to view the subject of the propaganda as less than human and consequently undeserving of the rights extended to humans. The subjects of such representation may be portrayed as lacking empathy, filthy and diseased, or possessing a unique proclivity towards criminality. Dehumanization through these tropes prepares the ground for the enactment of violence against members of the community in question. Whatever the context, the implications of such dehumanizing depictions remain the same: they normalize both physical and rhetorical violence against the targeted individual or community.

Our analysis revealed a distinct category of AI-generated images that strongly emphasized the exclusion and dehumanization of Indian Muslims. Comprising 291 posts, the narrative of

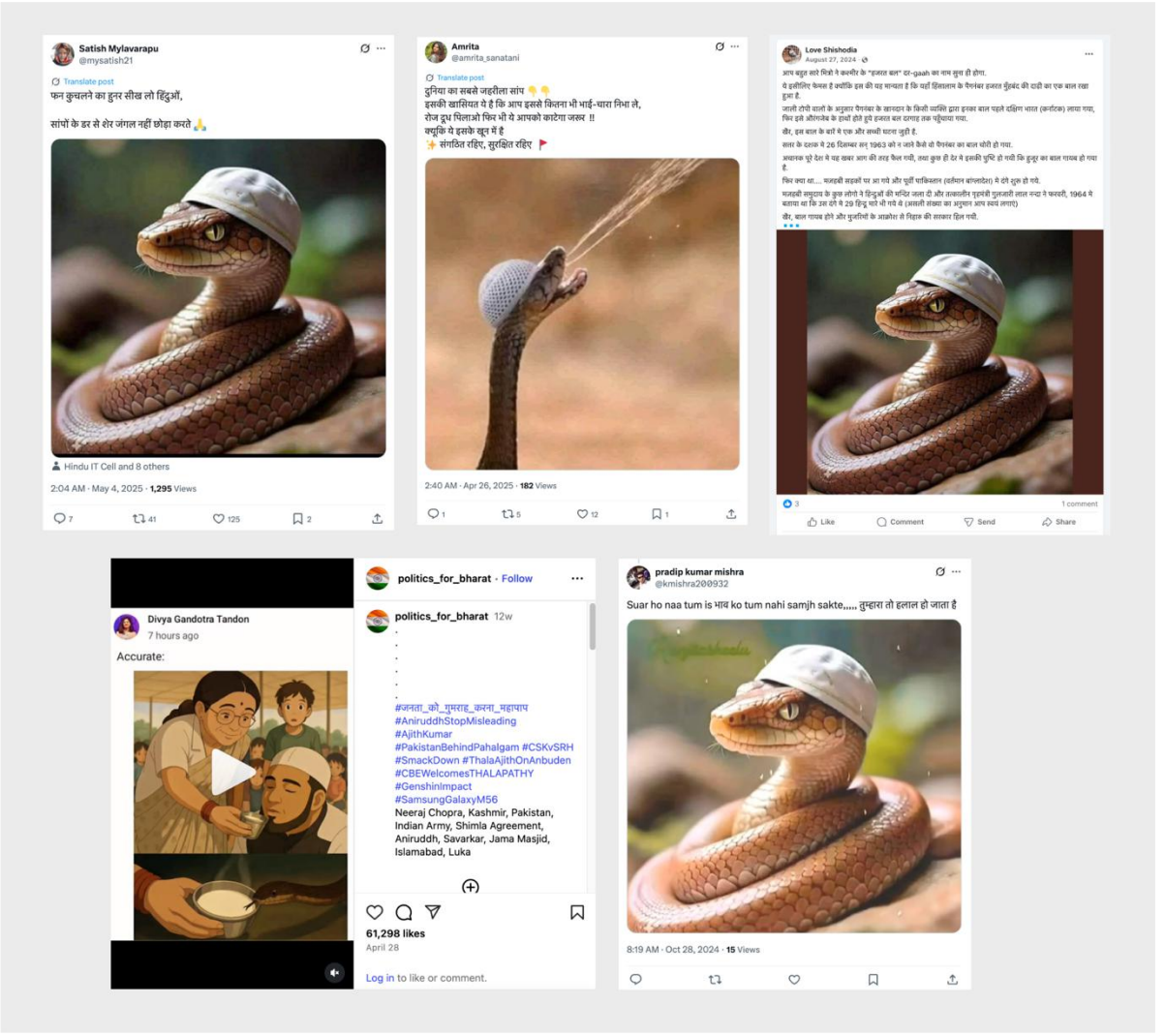
exclusion and dehumanization accumulated 5.79M views, 462.28K likes, 18.84K comments, and 130.94K shares, with a total engagement of 6.4M.

Below, we elaborate on the thematic subcategories through which dehumanization is expressed in AI-generated images targeting Muslims. Our analysis identifies two broad subcategories. The first, "Non-human Imagery: Animal Imagery as Religious Weaponization," includes both animal metaphors, such as snakes in skullcaps that frame Muslims as venomous and deceptive, and portrayals of Muslims as innately cruel toward animals, particularly cows and goats. The trope of Muslim cruelty toward animals also underscores the idea of animal slaughter in Muslim religious practices as barbaric. The second subcategory, "Framing Muslims as Threats: Crime, 'Infiltration,' and Terror," includes several tropes, including: the criminalization of the burqa (for example, the association of the burqa with crime or extremism); the depiction of Muslims as 'infiltrators' or illegal immigrants intent on demographic takeover; and the description of Muslims as sympathetic to terrorists, particularly in the aftermath of the Pahalgam terror attack of 2025. These subcategories and their constituent tropes provide the conceptual framework and context for the detailed visual and narrative analysis of posts in the larger category of exclusionary and dehumanizing rhetoric.

5.2.1 (A) NON-HUMAN IMAGERY: ANIMAL METAPHORS

Dehumanization through animal metaphors has been and continues to be a recurring device in hateful rhetoric worldwide. Likening a group to pests or predators erodes the boundary between symbolic and physical violence. Violence against the group portrayed as animals, typically dangerous creatures, pests, or vermin, becomes not only permissible but inevitable, reframed as a moral obligation and an urgent necessity to restore social hygiene. The use of animal metaphors to justify mass violence has deep historical roots. Tutsis were called "cockroaches"[46] by Hutu propagandists in Rwanda and Nazi propaganda likened Jews to rats[47]. In each case, the symbolic denigration of a group facilitated brutal and large-scale genocidal violence against them.

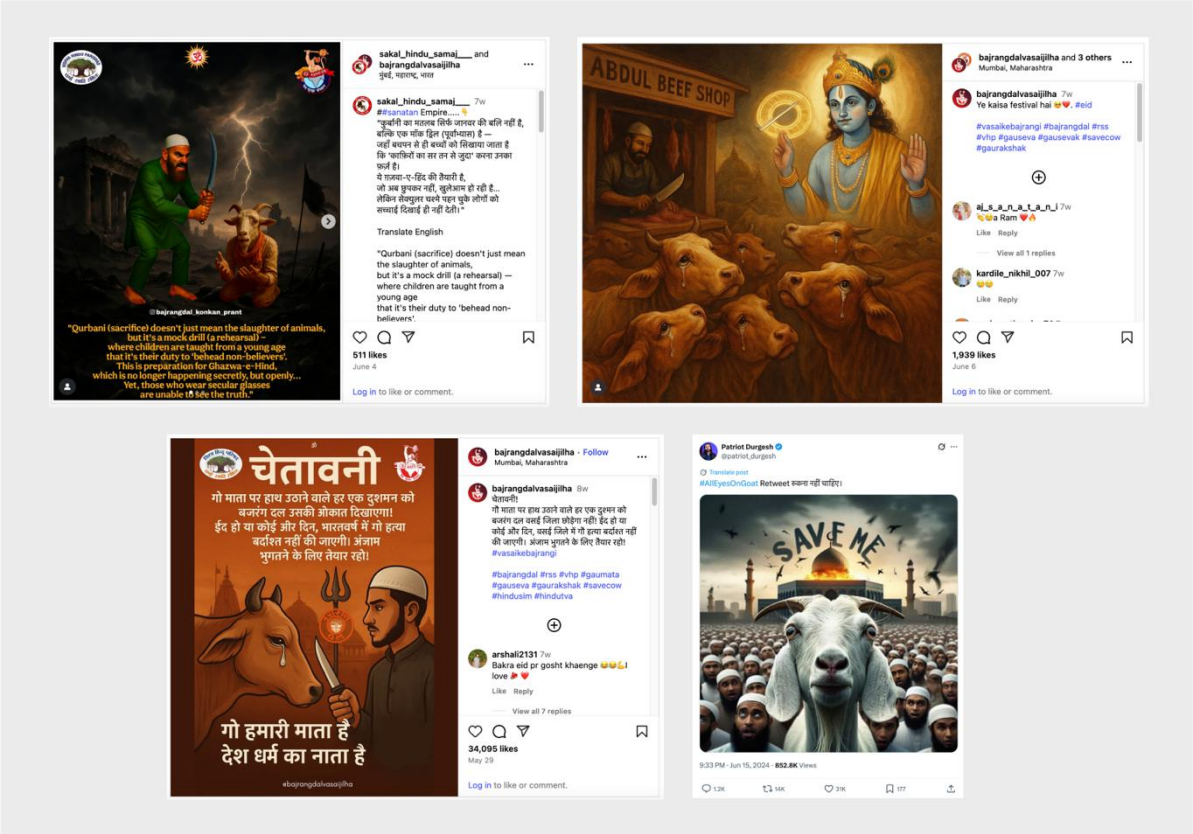
In Hindu nationalist rhetoric, Muslims are similarly and routinely portrayed as threats, pests, and enemies to be eliminated. We found AI-generated dehumanizing images that depict a snake wearing a skullcap. The snake is a potent symbol and a powerful metaphor for[48] depicting someone as less than human. The addition of the skullcap, an identifiable marker of Muslim identity, further cements the intended message. By likening Muslims to snakes, these AI-generated images frame them as deceptive, venomous, and deserving of extermination.



5.2.1 (B) NON-HUMAN IMAGERY: ANIMAL IMAGERY AS RELIGIOUS WEAPONIZATION

Another mode in which animal imagery is invoked emphasizes the idea that Muslims are cruel toward animals such as cows and goats, with special cruelty reserved for animals sacred to Hindus. This trope weaponizes Hindu cultural and religious symbols like the cow, thus designating the practice of religious slaughter by Muslims as barbaric. Such rhetoric has historically played a significant role in mobilizing mob violence against Muslims under the guise of "cow protection," a phenomenon for which social media has also been used by vigilante Hindu nationalist groups[49]. Cow slaughter prohibitions, enacted by Indian states, have also disproportionately targeted Muslims and other minorities. According to a Human Rights Watch report,[50] at least 44 people were killed in cow-related vigilante attacks between May 2015 and December 2018, 36 of whom were Muslims. Allegations of cow slaughter or beef consumption, often unverified or fabricated, have been used to justify brutal lynchings, public beatings, and social boycotts of Muslims. According to the IndiaSpend database,[51] more than 86 percent of victims in cow-related hate crimes between 2010 and 2017 have been Muslims.

A distinct thread of AI-generated images in our dataset centers on animals. Images of goats occur frequently in the archive of visual content, especially in the context of the Muslim religious festival of Eid al-Adha, the celebration of which entails the sacrifice of cattle. These images tend to anthropomorphize cattle, especially cows and goats, by depicting them with expressions of fear or pain. Viewers are encouraged to identify with the animal rather than with the human.



Through such visual rhetoric, the images construct Muslim ritual slaughter as an act of savagery and prepare the emotional ground for vigilante violence in the name of animal protection.

5.2.2 FRAMING MUSLIMS AS THREATS: CRIME, INFILTRATION, AND TERROR

The association of Muslims with a spectrum of nebulously defined threats has long been a staple of global Islamophobic discourse, including in the Indian context. Such framing is often vague but is nonetheless powerful, portraying Muslims as inherently untrustworthy, subversive, and dangerous to national unity and security. The idea of Muslim identity as threatening manifests itself in the criminalization of everyday Muslim practices and cultural expressions, which serves to legitimize social control and exclusion. Muslims are routinely depicted as a fifth column within the nation, permanently suspected of disloyalty. This subcategory examines the distinct ways in which such representations are disseminated through AI-generated images, which visually concretize and spread the narrative of the Indian Muslim as dangerous and untrustworthy.

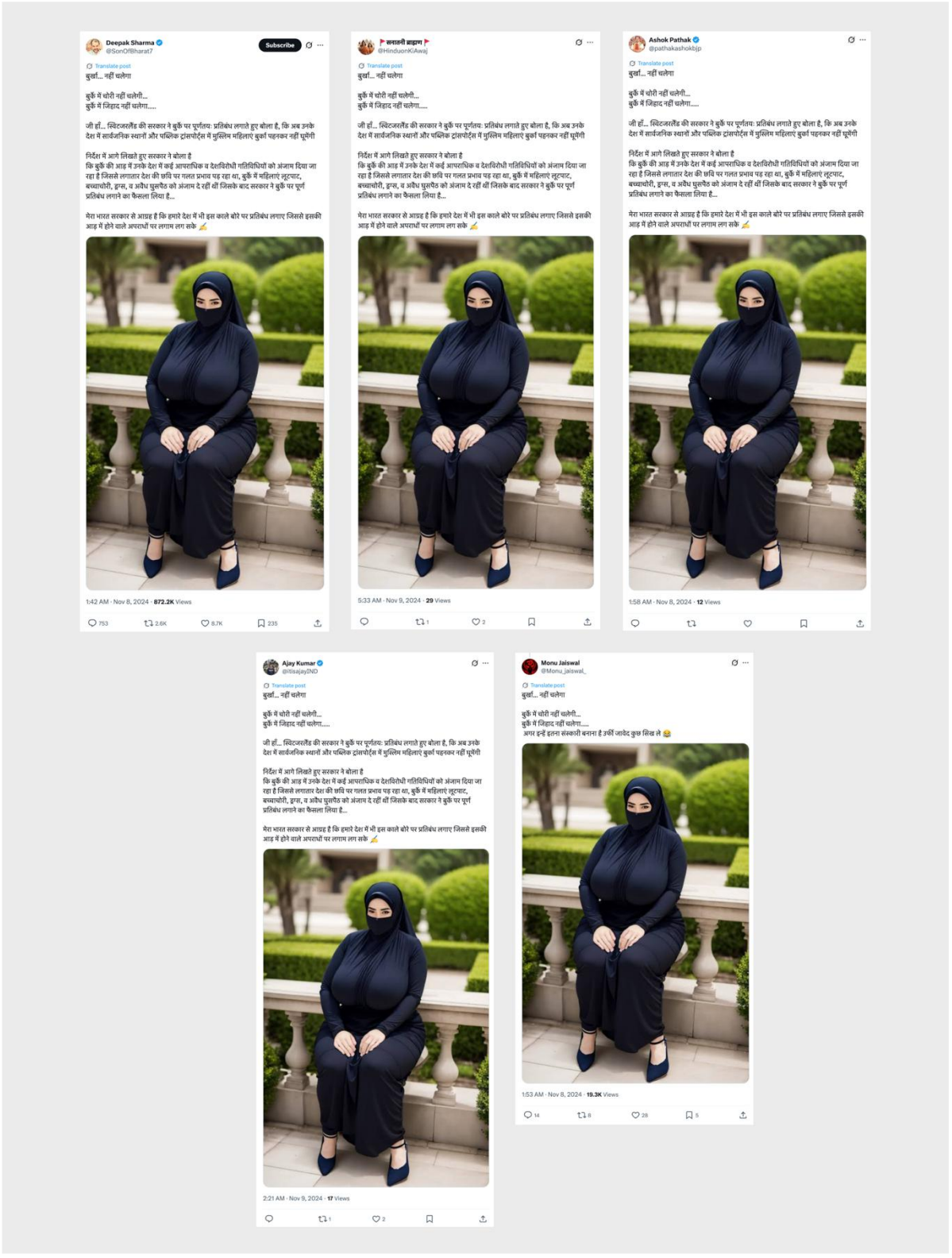
5.2.2 (A) CRIMINALIZATION OF THE BURQA

Beyond the routine sexualization of Muslim women, the burqa is a particular target of anti-Muslim discourse, reflecting a broader imperative that seeks to tarnish the entire Muslim community as criminal. AI-generated imagery that objectifies Muslim women through an obsessive focus on the burqa also casts them as security threats, aligning with a broader Islamophobic femonationalist[52] narrative. Femonationalism as a phenomenon started gaining special traction in France around 2010 and can also be seen in other countries that police Muslim women's dress. In India, the logic of femonationalism was clearly visible in the controversy over the 2022 hijab ban[53] in Karnataka state, which barred female Muslim students from classrooms on the grounds that their headscarves signified extremism. Such episodes illustrate how the hypervisibility imposed on Muslim women feeds a continuum of suspicion that degrades both their autonomy and the public perception of Islam. The term burqa/hijab in this context becomes a dog whistle in online rhetoric that targets Muslim identity at large.

In our analysis, we found several captions, written in Hindi, which directly stated that crimes committed under the cover of the burqa or, indeed, the very existence of the burqa, would no longer be tolerated. Some frequently voiced claims examples include the following statements:

- "बुर्खा... नहीं चलेगा" *("The burqa... will not be accepted.")*
- "बुर्के में चोरी नहीं चलेगी..." *("Theft in a burqa will not be tolerated...")*
- "बुर्के में जिहाद नहीं चलेगा....." *("Jihad in a burqa will not be tolerated...")*

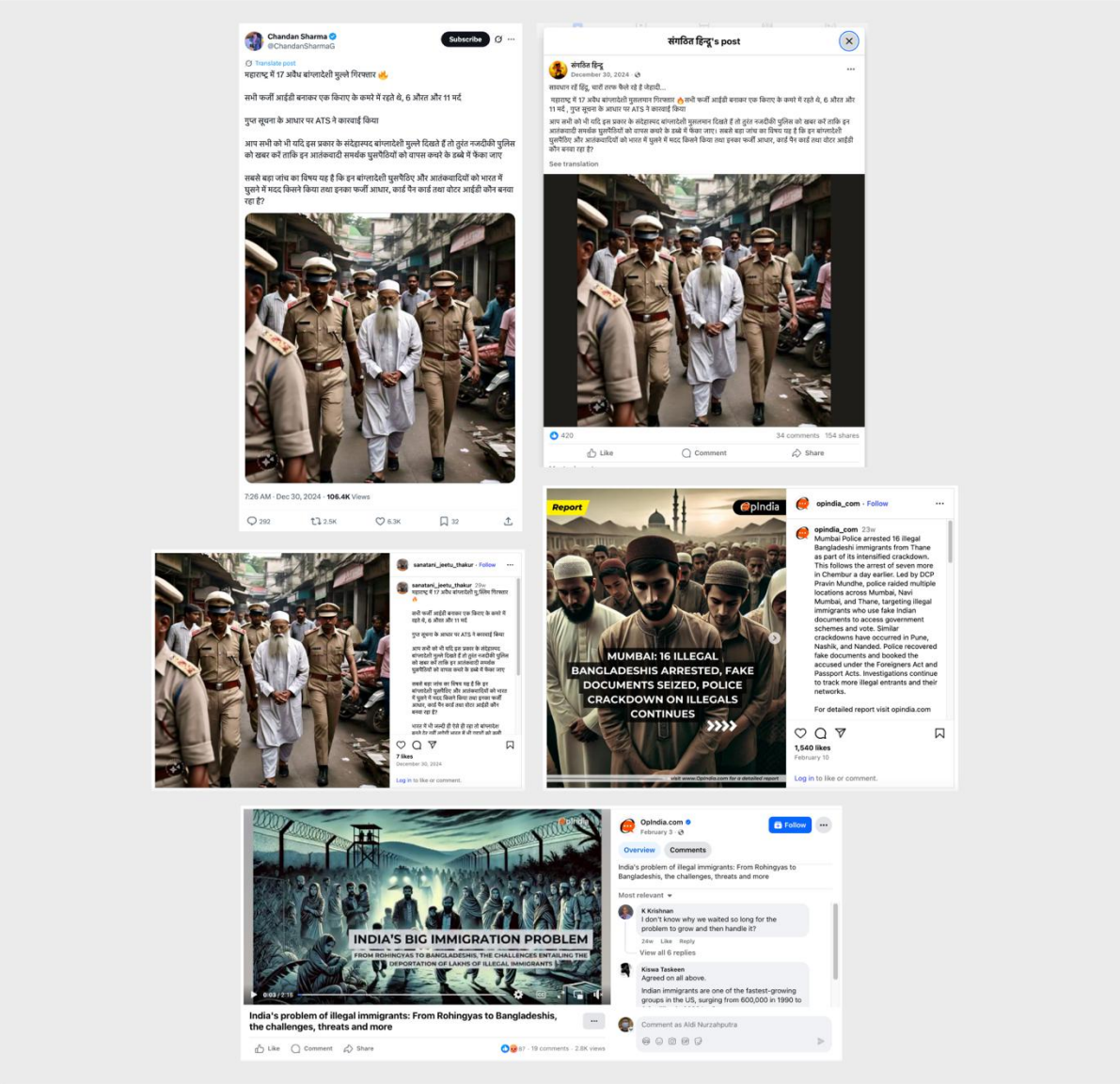
These captions conflate the burqa with criminal activity and religious extremism. The posts aim to ridicule and stigmatize Muslim women's clothing while also linking the attire of Muslims with threats to Indian national security.



5.2.2 (B) CASTING MUSLIMS AS "INFILTRATORS"

Hindu nationalist discourse in India has long branded Muslim citizens as "illegal infiltrators" who threaten the country's security, economy, and cultural identity. Hindu far-right politicians and media outlets have invoked and reinforced this trope for years. The trope is now

manifestly more visible with AI-generated imagery. AI-generated images showing Muslim men in skullcaps or women in burqas, visible religious markers that visually code them as both Muslim and foreign, dramatize an imagined demographic invasion. Captions with the images typically warn of fraudulent welfare claims by Muslims, forged national IDs used by them, or interfaith marriages conducted by Muslims with the alleged aim of seizing the property of Hindus.



The ruling Bharatiya Janata Party (BJP) has strategically mobilized[54] the same tropes in electoral contexts. For instance, during the Jharkhand Assembly elections in November 2024, senior BJP leaders, including Prime Minister Narendra Modi and Home Minister Amit Shah, repeatedly invoked the threat of alleged "Rohingya and Bangladeshi infiltrators" to stoke fears of demographic change and cultural erosion in various Indian regions. AI-generated images on these themes reinforce associations between Muslim identity and illegality, reinforcing xenophobic and Islamophobic stereotypes. In doing so, they play a powerful role in justifying exclusionary policies and normalizing discrimination against Muslims.

5.2.2 (C) THE AFTERMATH OF THE PAHALGAM ATTACK

Muslims have long been stigmatized as terrorists or sympathizers with terrorism, especially in the era following the War on Terror. The initial global focus on terrorism emerged in the early 2000s following the terrorist attacks of September 11, 2001. The stigmatization of Muslims as a community associated with violence and extremism has persisted and even intensified in many contexts since then. In India, Islamophobia in the aftermath of the War on Terror has reinforced and amplified Hindu right-wing anti-Muslim sentiment, leading to widespread suspicion of Muslims, discrimination against them, and the imposition and justification of harsh security measures targeting the community.

The power of the discourse of Muslims as violent extremists to mobilize hatred against the community was revealed after the tragic Pahalgam attack[55] on April 22, 2025, when four gunmen targeted tourists in Pahalgam in Jammu and Kashmir (J&K), killing 26 people in one of the region's deadliest terror attacks in recent years. India accused Pakistan of orchestrating the attack, a claim that Pakistan denied, further escalating already fraught tensions between the two countries. This event was quickly seized upon in the Indian public and political discourse to reinforce the narrative of Muslims as a security threat to the country.

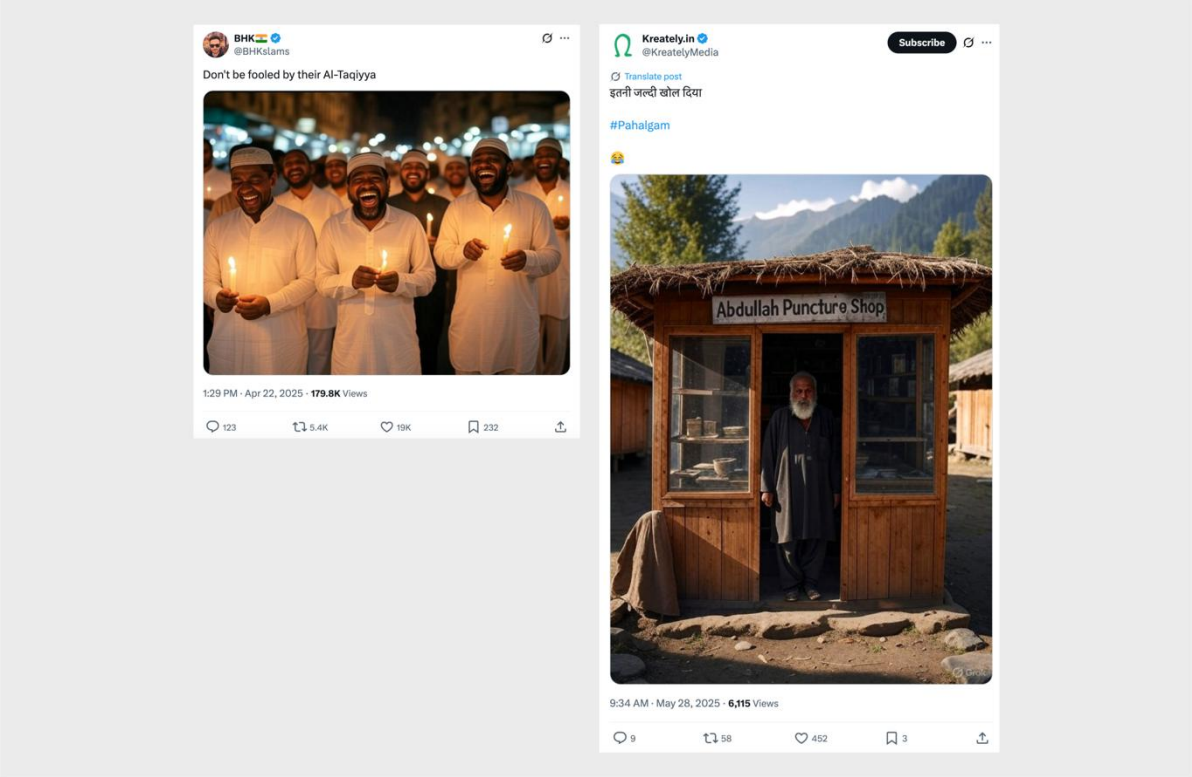
The Pahalgam terror attack was followed by an escalated conflict between India and Pakistan. The period following the attack and during and after the India-Pakistan conflict saw a massive spike in online misinformation, disinformation, and the use of AI[56] to promote hatred against Indian Muslims. Several accounts shared AI-generated visuals that suggested Muslims were celebrating the Pahalgam attack. Among them was a widely circulated post by verified X user @KreatlyMedia. The AI image purported to show Muslim men "laughing during candle marches" for the Pahalgam victims. The image was explicitly designed to provoke outrage, with the implication that Muslim participants were expressing joy during the memorialization of victims.

A post on X by The Jaipur Dialogues (@JaipurDialogues)[57] featured an AI-generated image of a Muslim man in traditional attire smoking against a background depicting a brawling mass of people. The image was captioned with references to India's internal political and social conflicts, such as Caste vs. Caste, Hindi vs. Tamil, and BJP vs. Congress. The caption reads, "Do you know the Kalma?" to refer to the Muslim belief. The image constructs Muslims as a people who primarily owe allegiance to their faith, with no national loyalty, as they nonchalantly observe India tearing itself apart. The image also reinforces the notion of Muslims as passive orchestrators of national discord.

The portrayal of Muslims as inappropriately jubilant in the wake of the violence against civilians and tourists in Pahalgam aims to simultaneously leverage public empathy for the victims and channel anger meant for the perpetrators of the attack toward all Indian Muslims.

These posts may have played a role in priming their audiences for expressing sectarian hostility towards Muslims and Kashmiris. In the aftermath of the Pahalgam attacks, 184 anti-Muslim and anti-Kashmiri hate crimes were recorded[58] between April 22 and May 8, 2025.

These patterns of AI-generated propaganda do not just spread hateful stereotypes against Muslims in India but systematically build a narrative that they are perpetual threats to the nation.



By combining narratives of criminality, infiltration, and callousness about the death of fellow Indians, AI-generated campaigns seek to influence the public perception of Muslims as outsiders undeserving of sympathy or rights.

5.3 SEXUALIZATION OF MUSLIM WOMEN

The theme of the sexual domination of Muslim women carries powerful symbolic weight in Hindu nationalist discourse. A clear reflection of the desire to subject the Muslim community to humiliation, and anchored in global Islamophobic rhetoric, it also reflects deep-rooted cultural nationalist anxieties. Events like the Gujarat 2002 pogrom[59] showed how sexual violence against Muslim women was used as a tool to annihilate the cultural identity of an entire community. Evidence confirms[60] that such sexual violence was normalized through leaflets by organizations like Vishwa Hindu Parishad and Bajrang Dal, both before and during the anti-Muslim riots. Videos and songs[61] in the recent past have also encouraged assaults of Muslim women during Hindu religious festivals.

The normalization of sexual violence has not remained confined to physical spaces, nor does it occur only during times of large-scale sectarian violence. It is also seen in the digital sphere, where the same dynamics of humiliation and control manifest themselves through the online targeting and harassment of Muslim women. Two examples are the Sulli Deals app controversy of 2021 and the Bulli Bai app episode of 2022. On both apps, photographs of hundreds of Muslim women, many of whom had vocally voiced their opposition to the BJP and the ideology of Hindu nationalism, were taken without consent from their social media

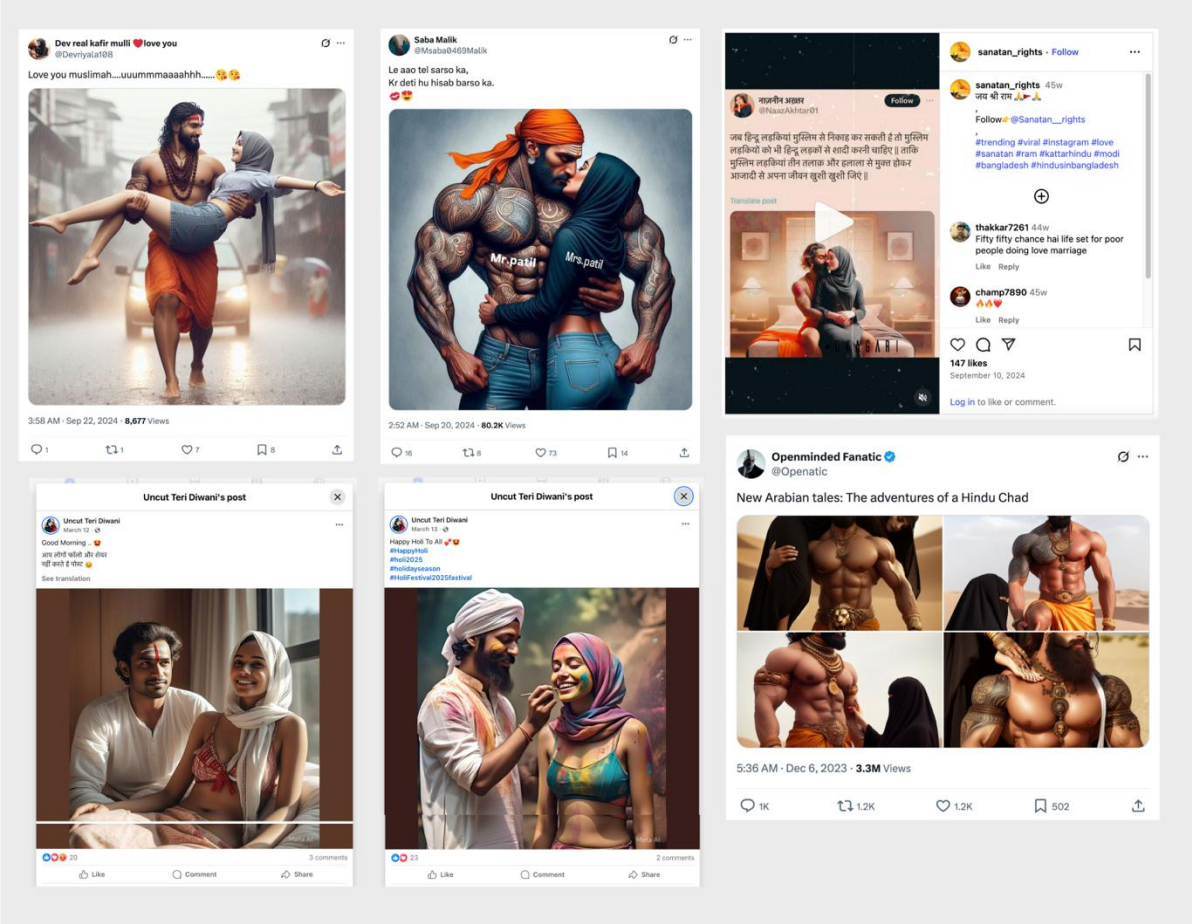
profiles and listed for “auctions” on a GitHub based app. The “winners” of the auction were encouraged to “claim” their prize by harassing the women on social media and sharing screen grabs of such harassment.

The sexualized depiction of Muslim women clearly emerges from a broader context of the normalization of violence against them. An investigation published by The Quint[62] in March 2025 revealed the widespread use of AI tools to create and distribute sexualized imagery of Muslim women. An earlier investigation[63] in 2021 had shown how similarly sexualized imagery (though not AI-generated) could be easily found on Twitter. The Quint investigation identified at least 250 Instagram pages posting AI-generated soft-porn images featuring women in hijabs or burqas in intimate positions with Hindu men. While these AI-generated or digitally altered images may not depict explicit sexual assaults, they nevertheless represent symbolic acts of violence. Unlike consensual intimate photographs or even explicit adult-themed sexual content, these images do not represent real people. Instead, they reflect the fantasies of those who create, share, and consume them; fantasies in which Muslim women are depicted without agency or consent. With 317 posts, this category had the highest total engagement of 6.7M in our dataset. The posts recorded 6.58M views, 89.82K likes, 11.54K comments, and 9.86K shares.

One AI-generated post shared by the handle @Openatic featured a hypersexualized fantasy of a Muslim woman in a burqa with a Hindu man and received tens of thousands of interactions (22.5K views, 23 replies, 66 RTs, 323 likes, and 38 bookmarks). The image is shared along with the caption “Why becoming a concubine is the best thing to happen to a muslimah after a war with the Hindus.”

We found numerous AI-generated images online that conform to the same pattern. Many depict Muslim women alongside Hindu-coded men, the latter often portrayed with exaggerated masculine features. The abaya, a long, loose-fitting cloak traditionally worn by some Muslim women as a form of modest dress, has become a visual shorthand in these images for Muslim identity.

In these examples, AI-generated imagery operates as both fantasy and threat. Muslim women are portrayed without consent or agency, cast as spoils of conquest in a narrative of domination. This framing not only sexualizes Muslim women but also dehumanizes Muslims, embedding misogyny within a broader Islamophobic project.



AI tools have lowered the barriers to producing and distributing such content, dramatically expanding both its scale and speed. What once required significant time and effort can now be fabricated within minutes and disseminated widely across platforms, making harassment more pervasive and far harder to contain. These hyper-realistic yet entirely fictional depictions of Muslim women strip them of consent and agency while functioning as symbolic acts of violence. They reinforce misogyny and Islamophobia by casting Muslim women as objects of conquest and domination, normalizing fantasies of subjugation, and inflicting real psychological and social harm on Muslim women in particular and the Muslim community at large.

5.4 AESTHETICIZATION OF VIOLENT IMAGERY

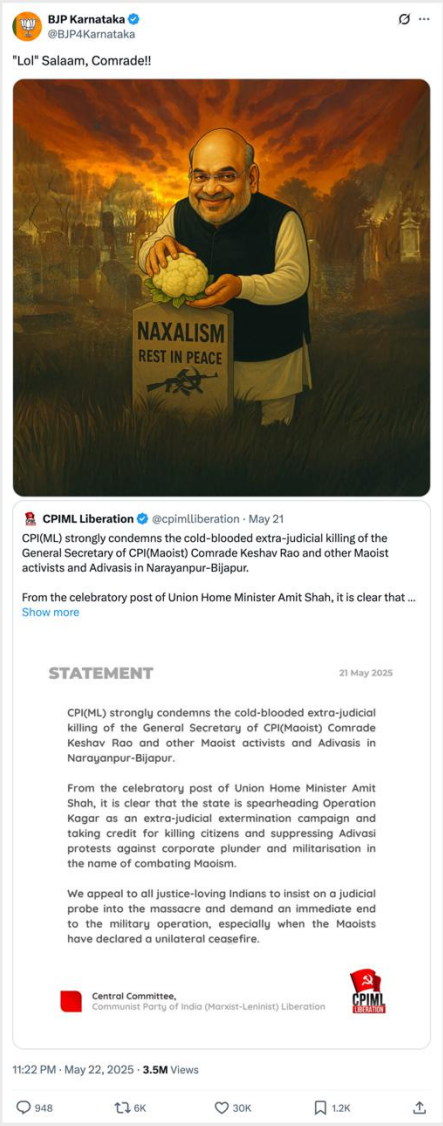
Recent developments in AI image generation, particularly with the launch of OpenAI's GPT-4o, have led to a rapid proliferation of high-quality, stylized images across social media platforms. Among the most popular of these are Ghibli-style[64] images, characterized by soft, dreamlike animations that mimic the aesthetic of the renowned Japanese animation studio. While used for whimsical or nostalgic content, this visual style has increasingly been appropriated to depict real-world incidents of Hindu nationalist violence.

The effective charge of such images, combined with their aesthetic familiarity—in this case, the Ghibli style—facilitates rapid circulation and heightened engagement among users online. Thus, it is not merely the use of AI but the convergence of affective manipulation, sectarian references, and platform dynamics that drives the visibility of such content. In our dataset, this category accounted for 134 posts, which garnered 5.15M views, 741K likes, 12.8K comments, and 93.5K shares, resulting in a total engagement of 6M.

The official X account of the BJP shared an AI-generated image featuring senior party leader L.K. Advani, accompanied by the slogan "Saugandh Ram ki khaate hain, mandir wahin banayenge" ("We swear by Lord Ram, we will build the temple at that very location").

In the 1980s and 1990s, this slogan was central to the Hindu nationalist mobilization that culminated in the destruction of the 16th-century Babri Masjid in Ayodhya on December 6, 1992, a disputed site, which according to Hindu nationalists had been built on the site of a temple marking the birthplace of the Hindu deity, Lord Ram. The demolition of the mosque marked a significant turning point in Indian politics, sparking nationwide riots in which thousands—mainly Muslims—were killed[65]. In the present context, the same slogan is visually reinforced through AI-generated symbols and images, embedding it within a digital landscape of Hindu nationalist memory of an imagined past of oppression at the hands of Muslim invaders and tyrants. The post received 417k views, 16K likes, 1.7k shares, and 331 comments on X.

The use of humor in anti-Muslim AI-generated content has been an effective tool[66] in evading the filters of content moderation and making hateful messages more widely accessible to online audiences. In May 2025, the official X account of the BJP in Karnataka shared an AI-generated image that sparked widespread controversy. The image depicted Home Minister Amit Shah holding a cauliflower[67].

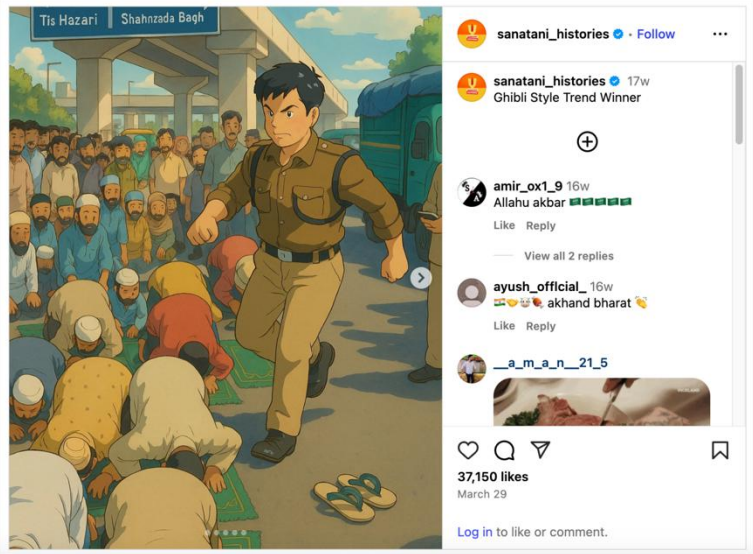


The image referenced the 1989 Bhagalpur anti-Muslim riots in Bihar, during which 110 Muslims were killed and their bodies buried in a cauliflower field. The trope was repurposed for recent police operations in the state of Chhattisgarh against a long-running left-wing insurgency, ongoing since the 1960s, known as the Naxalite movement. The post was a response to the condemnation of the alleged extra-judicial killing[68] of the members of the Naxalite group. It mocked both the victims of the killing and those who were expressing concern about the legality of the encounter. By invoking the imagery of past atrocities and mocking the victims of a present-day extrajudicial killing, the post sought to normalize both

mass majoritarian violence and state violence, presenting them as acceptable tactics in dealing with problem populations. The image received over 3.5 million views, 31k likes, 6k shares, and 951 comments.

This pattern of invoking and using past events for current political developments can also be seen in other examples in our dataset, in which AI-generated images are used to aestheticize violent milestones in India's troubled history of sectarian conflict. A widely circulated Ghibli-style image posted on Instagram depicted the demolition of the Babri Masjid in 1992. The image portrays thousands of Hindu nationalist supporters scaling the domes of the mosque, with many of them waving saffron flags. Posted by @ThePulkitSBisht on Instagram with the caption "The Watershed Moment 🚩," the image received over 4 million views, 21k likes, and thousands of shares, transforming a moment of violent social and political rupture into a visually aestheticized and celebratory spectacle. Repackaging a traumatic historical event for Muslims as digital art allows AI-generated imagery to facilitate the troubling reframing of public memory, glorifying violence while erasing the suffering caused by the event.

This visual strategy has also been used to represent recent incidents of state violence against Muslims. A widely circulated Ghibli-style image on Instagram, for instance, portrays a Delhi police officer kicking Muslim worshippers during namaz (prayer), an AI recreation of a real incident from March 2024[69]. Rendered through a soft, animated lens, the brutality is visually sanitized, turning an act of violence into something deceptively palatable. Such aesthetic manipulation raises urgent questions about how AI-generated images blur the line between fiction and reality, desensitize viewers to brutality, and shape public opinion about the legitimacy of violence against minorities.

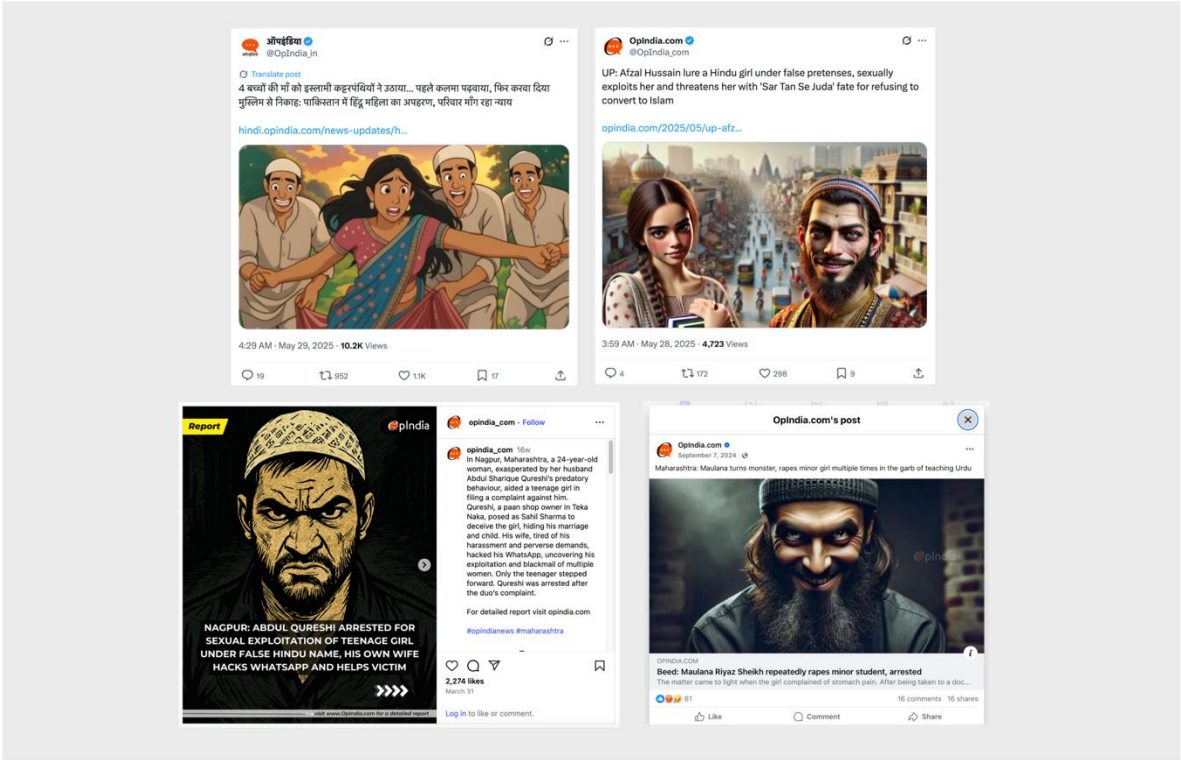


6. HINDU FAR-RIGHT MEDIA AND THE WEAPONIZATION OF AI

Generative AI content has quickly proliferated across Indian newsrooms, with many media houses using AI-generated images and even AI newsreaders or anchors[70]. Across India's Hindu far-right media ecosystem, AI-generated images have emerged as a powerful tool for amplifying Islamophobic narratives. Key Hindu far-right platforms such as OpIndia, Sudarshan News, Satyaagrah, and affiliated actors like Nupur J. Sharma and the Hinduphobia Tracker, use AI-generated images to frame Muslims as deceitful, a socially malignant presence, and violent.

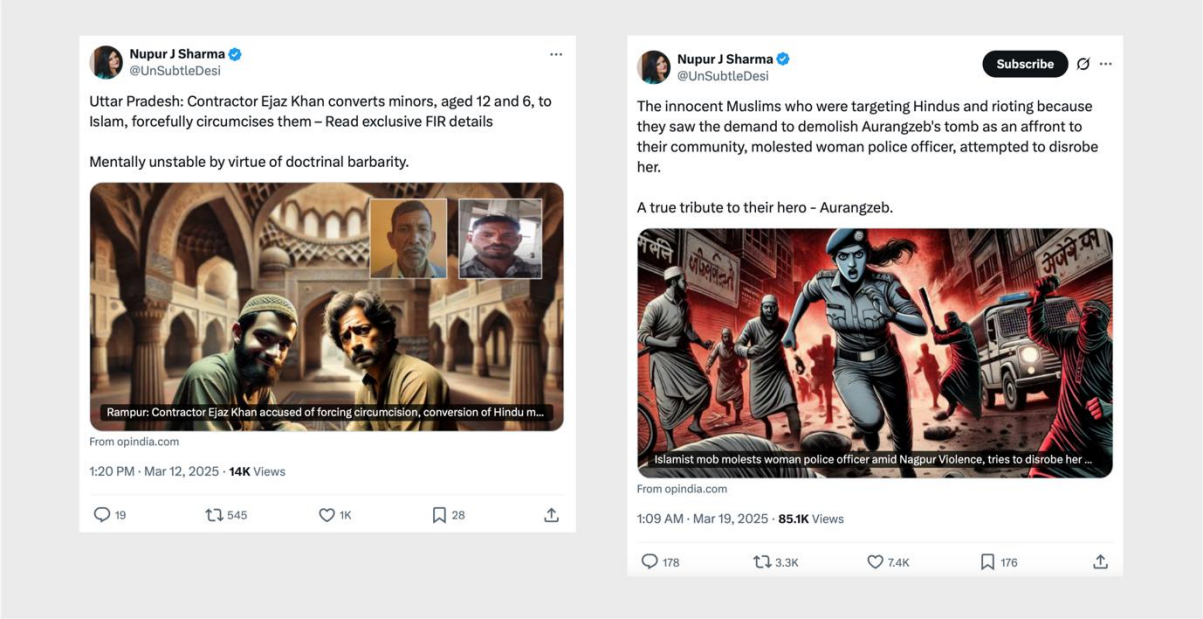
6.1 THE OPINDIA NETWORK

The OpIndia network comprises nine accounts operating across both its Hindi and English outlets. This network includes the Hinduphobia Tracker project, which is directly affiliated with OpIndia through shared leadership and organizational ties. Nupur Sharma, the editor of OpIndia, is a key figure involved in content production across this network. Collectively, these accounts were responsible for 262 posts in our dataset, generating approximately 1.1 million engagements, including 601.3K views and 462.8K likes. The platform operates in both English and Hindi and has played a major role in popularizing AI-generated images that depict Muslims as violent, deceptive, or a socially threatening force. Our dataset reveals that OpIndia has shared multiple AI-generated images yoked to narratives about forced conversions, abductions, and religious coercion by Muslims.

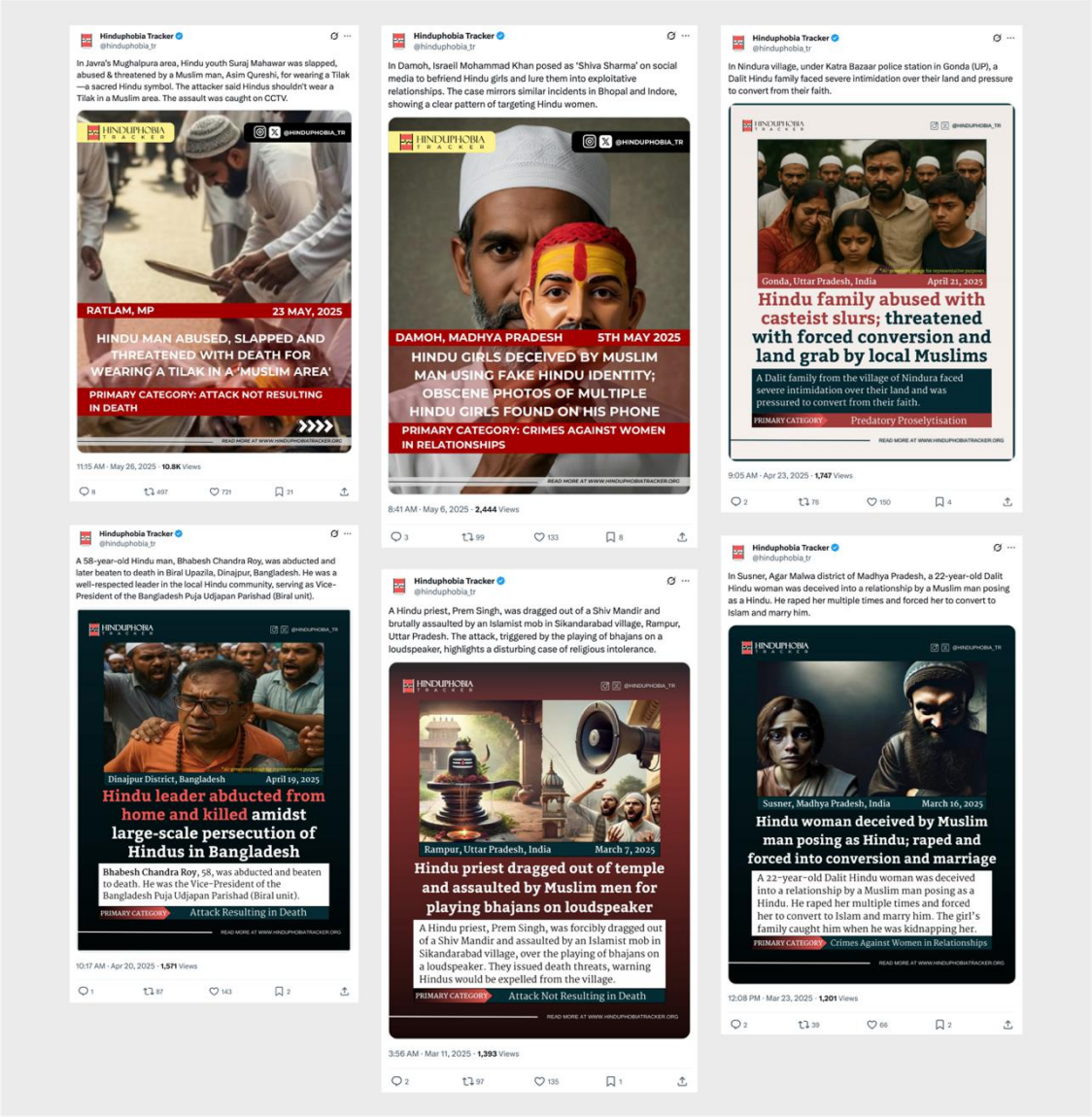


The outlet's editorial leadership and ownership structure reveal strong ideological proximity to the ruling Bharatiya Janata Party. Director Ashok Kumar Gupta and editor-in-chief Nupur J. Sharma have longstanding public affiliations with the BJP[71], with Sharma playing a dual role in the Hindu right-wing ecosystem as both media executive and an independent influencer on social media.

Sharma has repeatedly shared AI-generated images advancing claims about Muslims related to the circumcision of minors, conversion of non-Muslims through deceit, and theological extremism. In one example, an image shows two men with caricatured features inside a mosque, accompanied by a caption describing them as "mentally unstable by virtue of doctrinal barbarity," an assertion that debases both the individuals and the faith they represent.



The Hinduphobia Tracker, which presents itself as a civil documentation initiative, operates under the Gavishti Foundation, led by Sharma and Rahul Roushan, CEO of OpIndia. This arrangement creates a closed-loop system of production, amplification, and ideological reinforcement. AI-generated images circulated by the Hinduphobia Tracker follow the same tropes as those produced and spread by the OpIndia network. Posts from the Hinduphobia Tracker routinely featured AI-generated images showing Muslim men involved in violence, desecration of religious symbols or structures of other communities, or criminal acts. For example, in the wake of the Pahalgam attack of 2025, the Hinduphobia Tracker shared AI-generated images promoting the narrative that Indian Muslims sympathized with the attackers. One of these posts, shared across X[72], Instagram[73] and Facebook[74], featured an AI-generated image of a man clearly coded as Muslim, through the presence of his skullcap, allegedly celebrating the attack. The tone and presentation of such posts reflect the editorial slant of the OpIndia ecosystem, which is marked by a convergence of personnel, content, and ideological unity across its component platforms.



6.2 SUDARSHAN NEWS

Sudarshan News, a Hindi-language far-right television channel, routinely uses AI-generated thumbnails and segment backdrops that present fabricated scenarios as unimpeachable visual evidence. Sudarshan News played a prominent role in promoting the Rail Jihad conspiracy narrative described above in the report. Channel anchors Suresh and Pravesh Chavhanke, as well as nine other accounts associated with the channel, produced 187 posts characterized by hateful anti-Muslim content and the use of AI. These posts generated about 569.7K total engagements, with 488.7K views and 68.6K likes.

One incident referenced by the channel was the tragic accident involving the Pushpak Express in Jalgaon, Maharashtra, in January 2025[75]. Around 4 PM, passengers on the 12533 Lucknow–Mumbai Pushpak Express noticed sparks inside one of the coaches and suspected a fire. Passengers pulled the emergency chain, which halted the train near Pachora in Jalgaon district. The sparks were later attributed to a hot axle or brake binding, not a fire. In the

ensuing panic, several passengers jumped off the halted train onto an adjacent track, unaware that another train was approaching. They were struck by the oncoming train while trying to flee. The mishap resulted in at least eleven deaths.

Sudarshan News portrayed the tragedy as planned and targeted sabotage by Muslims. In a post circulated on X, Facebook, and Instagram, the channel used an AI-generated image of a Muslim man, identifiable through a beard, skullcap, and green attire associated with Islam, holding an axe with a train in the background. The image was not presented as symbolic or illustrative but framed as concrete visual proof of the Rail Jihad theory. Captions accompanying the post warned of a coordinated plan by "jihadi" actors to attack railway infrastructure, even though no investigative evidence or law enforcement agency suggested such a motive.

Following the Pahalgam terror attack, the channel circulated AI-generated images calling for the economic boycott of Muslims and portraying the community as internal enemies of the Indian nation. These were accompanied by jingoistic slogans such as अब भारत के अंदर के पाकिस्तान को ठोको मोदी जी ("Modi Ji, strike down the Pakistan that exists within India"). Such images redirect public anger over the Pahalgam attack toward Indian Muslims, layering fiction onto tragedy and making unproven conspiracies appear not only plausible but visually self-evident.

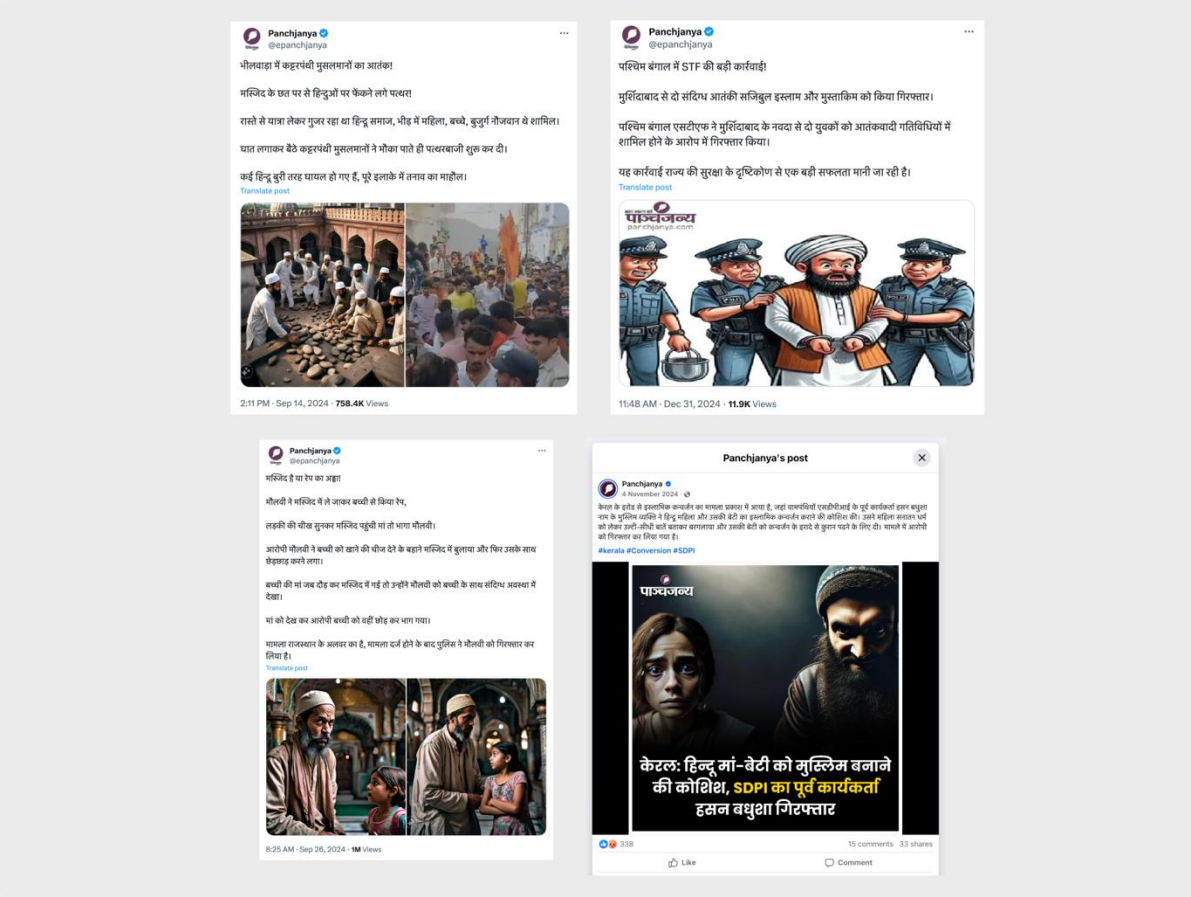
Sudarshan News frequently deploys AI-generated images to reinforce the Love Jihad, Thook Jihad (spit jihad), and other jihad-themed conspiracy theories, which frame Muslims as a societal threat. Many of these visuals depict Muslim men with snake tongues protruding from their mouths, combining Muslim-coded identity with non-human monstrosity. During the Mahakumbh, a major Hindu festival in January 2025, the channel circulated AI-generated images portraying Muslims as demonic figures with horns, promoting the false narrative that Muslims were conspiring to disrupt the event.

Across these posts, captions play a crucial role in shaping audience interpretation. Sudarshan News repeatedly uses direct calls to action such as “Jaago” (“Wake up”) and “Jaago Hinduon” (“Wake up, Hindus”). These commands function as emotional triggers, urging Hindus to recognize a clear threat and mobilize collectively against Muslims. The framing equates passivity with betrayal, while action, whether electoral, discursive, or violent, is cast as patriotic and religious duty. Paired with AI-generated visuals of threatening-looking Muslims and fabricated acts of confrontation, the captions provide a narrative key that orients viewers to feel both fear and hostility.

Sudarshan’s strategy of using AI-generated visuals circumvents the need for documentation. Instead of verified evidence, the channel constructs an imagined reality aligned with its editorial agenda. Segments first aired on television are repackaged with fabricated thumbnails and disseminated widely on social media. The effect is cumulative: fabricated narratives layered over real-life tragedies blur the line between fact and fiction. As a result, baseless conspiracies acquire an unsettling veneer of credibility, appearing not just plausible but indisputable, encouraging fear of Muslims and deepening mistrust and hostility toward them.

6.3 PANCHAJANYA

Panchjanya, a Hindi-language weekly magazine published by the Rashtriya Swayamsevak Sangh (RSS), the ideological parent of the ruling BJP, has repeatedly used AI-generated images to frame Muslims as violent and treasonous. The magazine's single X account, though producing only 29 posts, drew more than 5.2 million views and approximately 5.3 million total engagements. In one post, Panchjanya shared an AI-generated image of Muslim men gathering stones on the terrace of a mosque, insinuating that they were preparing to attack Hindus during Ganpati festival celebrations. The image, however, was fabricated and bore no connection to any documented incident reports.



The repeated use of AI-generated images by Panchjanya not only reinforces staple themes of anti-Muslim discourse but also influences how audiences perceive and understand Indian Muslims as threats, invaders, or cultural saboteurs. Though these Hindu nationalist media platforms vary in format, encompassing weekly magazines, digital news, reels, and thumbnails, the underlying narrative goal of the content that they produce is the same: to create believable fictions that solidify anti-Muslim prejudice, normalize hostility, and mobilize audiences toward collective action.

7. FAILURE TO ENFORCE HARMFUL CONTENT POLICIES AND COMMUNITY GUIDELINES

To evaluate the effectiveness of reporting tools on social media platforms, we flagged 187 harmful posts featuring AI-generated images that were in clear breach of the platforms' own community standards. Of these, 108 were on X, 37 on Facebook, and 42 on Instagram. The posts were reported under the platforms' designated categories: on X as "Hate," "Abuse & Harassment," and "Violent Speech," and on Facebook and Instagram as "Hate Speech" and "Calls for Violence."

Despite violating platform policies, only one of the flagged posts on X had been removed at the time of writing this report. The remaining 186 posts remained online, underscoring the platform's failure to enforce its stated community standards.

The inaction is particularly alarming in the context of AI-generated content. Unlike traditional forms of harmful content, AI tools allow hate actors to rapidly produce slopaganda and related content. If left unchecked, it can saturate the information environment, spread across networks, and normalizes visual propaganda that would once have required significant resources to manufacture.

The failure to act on reported AI-generated hate means platforms are not merely ignoring violations but accelerating new forms of digital harm. This exposes a dangerous gap between the rapid evolution of AI technologies and the sluggishness of current enforcement models. To close this gap, platforms must invest in AI-aware moderation systems, strengthen synthetic media detection, and apply their rules with consistency.

8. CONCLUSION

This report documents the emerging role of AI-generated imagery in the production and circulation of anti-Muslim hate in India's digital sphere. By analyzing 1,326 posts across X, Facebook, and Instagram that were produced between January 2024 and April 2025, we examine how AI-generated images are being weaponized to dehumanize Muslims, propagate conspiracy theories, aestheticize violence, and normalize misogynistic and Islamophobic narratives.

Generative AI now enables the amplification of old tropes through new visual grammars: Muslims depicted as violent, immoral, or deceitful; Muslim women objectified and hypersexualized; and Islamic symbols reframed as threats to national integrity. This is not a departure from past practices of photoshopping or crude meme-making[76]. Rather, it represents the acceleration of such practices, through the production content with greater credibility and precision and at scale. At the core of this transformation lies a convergence between individual users, far-right media outlets, platform infrastructures, and generative AI tools, which together create a feedback loop that entrenches synthetic hate and enables it to evolve in new forms.

The role of Hindu nationalist media actors such as OpIndia, Panchjanya, and Sudarshan News is central to this transformed apparatus of the production of online hate. With their reach and credibility as established media brands among a large section of Indians in present-day India, they embed AI-generated hate into mainstream discourse, where it circulates widely, receives millions of engagements, and is legitimized further through recirculation by ordinary users. Platforms have failed to address these harms. Their moderation policies remain ineffective. And, even if policies are enforced, the viral spread and pervasiveness of hateful content points to systemic gaps in policy enforcement.

The risks are profound. For Indian Muslims, this visual propaganda deepens the climate of fear, humiliation, and social exclusion that is becoming an entrenched feature of Indian society. Muslim women face a distinct and gendered threat, as AI imagery renders them hypervisible while normalizing their sexual subjugation. In terms of the implications for Indian society, such hateful content accelerates the dehumanization of a religious minority, corrodes constitutional protections for minorities, and weakens the foundations of Indian democratic institutions.

The implications of AI-generated hate extend beyond India. Social media companies risk losing user trust as hateful visual content dominates feeds. Generative AI companies risk their tools being irreversibly associated with low-quality, abusive, and harmful outputs. With the introduction of low-cost ChatGPT subscriptions, the Indian internet space is likely to see a sharp surge in the creation, circulation, and consumption of AI-generated visual content.

Coupled with an already flourishing Hindu far-right online ecosystem dedicated to producing anti-Muslim hate, the conditions are in place for an exponential rise in such material.

Addressing this crisis requires immediate, coordinated action. AI companies must integrate harm-prevention mechanisms at the design level to limit the weaponization of their tools. Social media platforms must significantly improve their detection, labeling, and moderation of synthetic hateful content, with greater transparency and accountability in enforcement. Without intervention, AI-generated hate will harden into the architecture of India's digital sphere.

9. RECOMMENDATIONS

The findings of this report make it clear that AI text-to-image technology has been weaponized within India's volatile digital ecosystem for the production and circulation of harmful content targeting Muslims. In light of the patterns identified in this report, a multi-pronged response is merited. In particular, we emphasize that nearly all stakeholders in the digital ecosystem, including platforms, civil society organisations, and lawmakers, must take a proactive role in addressing the harms.

1. **Targeted legal provisions for AI-generated content:** India's existing legal framework, specifically the Information Technology Act, 2000[77], is not adequately equipped to address the unique challenges of Generative Artificial Intelligence (GAI), specifically text-to-image and video tools. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021[78] expanded intermediary responsibilities and introduced due diligence requirements for content removal, but do not provide a framework for identifying or regulating the production of AI-generated images and videos through GAI technology. The UNESCO Guidelines for the Governance of Digital Platforms, 2023[79] provides a basic outline of a rights-compatible approach that can be followed. Clear and auditable transparency duties must be placed directly on model providers, such as requiring comprehensive model and risk documentation. Specified high risk uses such as Non-Consensual Intimate Images (NCII) and biometric imagery can be subjected to greater disclosure requirements involving social media platforms.
2. **Specifying adjudicatory authorities:** The proposed regulation can also specify adjudicatory authorities that prevent the shift of proactive detection responsibilities to intermediaries. These could draw from the out-of-court dispute settlement process offered by Article 21 of the EU DSA[80]. This would allow both user rights to the freedom of expression be preserved and platforms ability to enforce community standards.
3. **Developer safety and transparency:** AI developers and model hosts must share responsibilities in ensuring transparency regarding AI-generated imagery. In addition to clearly communicating that hateful use of such imagery is prohibited, developers should implement and continually refine strong detection, reporting, and moderation systems to identify misuse in real time.
4. **Provenance-first defaults and abuse prevention in AI tools:** AI developers and model hosts must build in safety and transparency by default. This means rather than react to the violations of user guidelines when reported, platforms should actively develop their features in a manner that it is difficult to abuse them. Every image output should carry embedded provenance metadata, ensuring that synthetic content can be reliably traced back to its source.

5. **Standards for AI use in media:** Media responsibility is a crucial pillar of maintaining a reliable information environment. India's news media regulatory bodies, including the News Broadcasting & Digital Standards Authority (NBDSA), should expand their jurisdiction to explicitly cover AI-generated images and videos used in news content. Similarly, the Press Council of India should issue guidelines requiring clear disclosure whenever AI-generated images appear in news media and develop voluntary codes of practice to govern the ethical use of AI-generated content.
6. **Establishing Open-Source Research Databases:** National regulatory authorities and social media companies must work to establish a coordinated civil society consortium enabling digital rights NGOs, fact-checking collectives, and academic labs to act collectively as an independent watchdog about AI-generated hate imagery. The consortium could build and maintain open, regularly updated datasets of synthetic hate images for research and model-testing.
7. **Algorithmic Transparency and Independent Audits:** Tech platforms must assess and disclose how recommendation and ranking systems amplify AI-generated hateful content. Moreover, they should enlist the support of independent auditors to test amplification, downranking efficacy, repeat-offender handling, and cross-language performance, with public summaries and corrective action plans. Furthermore, platforms should publish standardized risk metrics, such as time-to-label and take down, recommendation rates for detected synthetic hate, and recidivism, disaggregated by language and state.
8. **Cross-platform Early Warning and Joint Response:** Platforms should convene a privacy-preserving consortium of platforms, fact-checkers, and civil society organizations to detect and disrupt synthetic hate campaigns[1]. Participants should exchange hashes, provenance signals, and campaign indicators to enable synchronized throttling, labeling, and lawful content takedowns.
9. **Friction for Viral Synthetic Hate:** Platforms should introduce graduated "circuit breakers" protocols to thwart the spread of harmful synthetic content. Platforms must consider stricter rate limits, demonetization, and link-out suppression to repeat offender accounts, with delivery reinstated only after verified remediation.

[1] Coordinated Inauthentic Behavior (CIB), which refers to synthetic campaigns, is a term used by social media platforms, notably Meta, to describe the use of multiple accounts, pages, or groups working together to mislead users about their identity, purpose, or origin, often to manipulate public debate or amplify particular narratives.

10. ENDNOTES

1. NVIDIA. "Generative AI – What Is It and How Does It Work?" NVIDIA n.d. https://www.nvidia.com/en-us/glossary/generative-ai/.
2. Press, Gil. 2022. "What Happened to AI in 2022?" Forbes, December 30, 2022. https://www.forbes.com/sites/gilpress/2022/12/30/what-happened-to-ai-in-2022/.
3. Center for the Study of Organized Hate. n.d. "Resources: Reports." Center for the Study of Organized Hate. https://www.csohate.org/resources/reports/
4. Butts, Dylan. 2025. "OpenAI Chases Growth in India With Its Cheapest ChatGPT Plan, Costing $4.6 a Month." CNBC. August 20, 2025. https://www.cnbc.com/2025/08/19/openai-chases-growth-india-with-cheapest-plan-at-chatgpt-plan-at-399-rupees-global-rollout.html.
5. Levy, Steven. 2015. "Inside Deep Dreams: How Google Made Its Computers Go Crazy." WIRED, December 11, 2015. https://www.wired.com/2015/12/inside-deep-dreams-how-google-made-its-computers-go-crazy/.
6. OpenAI. 2024. "DALL·E Now Available Without Waitlist." 2024. OpenAI. March 13, 2024. https://openai.com/index/dall-e-now-available-without-waitlist/.
7. Press Information Bureau. 2024. "Universal Connectivity and Digital India Initiatives Reaching to All Areas, Including Tier-2/3 Cities and Villages." Press Information Bureau. August 2, 2024. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2040566.
8. LocalCircles "1 In 2 Indian Internet Users Are Already Using AI Platforms." n.d. LocalCircles. https://www.localcircles.com/a/press/page/india-ai-survey.
9. Biswas, Soutik. 2024. "'Invisible in Our Own Country': Being Muslim in Modi's India." BBC. April 28, 2024. https://www.bbc.com/news/world-asia-india-68498675.
10. Regeni, Petra and Claudia Wallner. 2025. "Online Hate Speech and Discrimination in the Age of AI" Royal United Services Institute, June 4, 2025. https://static.rusi.org/online-hate-speech-and-discrimination-in-the-age-of-ai-june-2025.pdf
11. Quinn, Ben, and Dan Milmo. 2024. "How The Far Right Is Weaponising AI-generated Content in Europe." The Guardian, November 26, 2024. https://www.theguardian.com/technology/2024/nov/26/far-right-weaponising-ai-generated-content-europe.
12. Brown, Mark. 2025. "Inquests Into Deaths of Southport Stabbing Victims Opened and Adjourned." The Guardian, January 20, 2025. https://www.theguardian.com/uk-news/article/2024/aug/07/inquests-deaths-southport-stabbing-victims-opened-adjourned.
13. Tondo, Lorenzo. 2025. "Italian Opposition File Complaint Over Far-right Party's Use of 'Racist' AI Images." The Guardian, April 18, 2025. https://www.theguardian.com/technology/2025/apr/18/italian-opposition-complaint-far-right-matteo-salvini-lega-racist-ai-images.
14. Bertoldini, Lucia. 2025. "How The European Far-right Is Using AI-generated Content to Engage Voters" European Digital Media Observatory, May 6, 2025. https://edmo.eu/publications/how-the-european-far-right-is-using-ai-generated-content-to-engage-voters/.
15. Skelly, Susan. 2025. "'Slopaganda' and Its Potential to Upend Elections on a Knife Edge." The Lighthouse. April 24, 2025. https://lighthouse.mq.edu.au/article/april-2025/slopaganda-and-its-potential-to-upend-elections-on-a-knife-edge.
16. Koebler, Jason. 2025. "AI Slop Is a Brute Force Attack on the Algorithms That Control Reality." 404 Media. March 28, 2025. https://www.404media.co/ai-slop-is-a-brute-force-attack-on-the-algorithms-that-control-reality/.
17. Jugendschutz. 2025. "KI-Bildertrends: Wie Extremist:Innen Aktuelle Filter Und Stile Für Ihre Zwecke Einsetzen." Jugendschutz. April 2025. https://www.jugendschutz.net/themen/politischer-extremismus/artikel/ki-bildertrends-wie-extremistinnen-aktuelle-filter-und-stile-fuer-ihre-zwecke-einsetzen.
18. eSafety Commissioner, 2024 "eSafety Urges Schools to Report Deepfakes as Numbers Double | eSafety Commissioner." eSafety Commissioner. June 27, 2024. https://www.esafety.gov.au/newsroom/media-releases/esafety-urges-schools-to-report-deepfakes-as-numbers-double
19. Center for the Study of Organized Hate. 2025. "Report 2024: Hate Speech Events in India" Center for the Study of Organized Hate. February 10, 2025. https://www.csohate.org/2025/02/10/hate-speech-report-india-2024/

20. Center for the Study of Organized Hate. 2024. "Streaming Violence: How Instagram Fuels Cow Vigilantism in India" Center for the Study of Organized Hate. November 19, 2024. https://www.csohate.org/2024/11/19/instagram-fuels-cow-vigilantism-in-india-report/
21. Chaudhuri, Pooja and Yong Xiong. 2023. "YouTube hosts Hindutva pop videos that violate its hate policies, auto-generates more" Columbia Journalism Review. March 8, 2023. https://www.cjr.org/tow_center/youtube-allows-hindutva-pop-videos-that-violate-its-hate-policies-auto-generates-more.php
22. Zakrzewski, Cat, Gerrit De Vynck, Niha Masih, and Shibani Mahtani. 2023. "How Facebook Neglected the Rest of the World, Fueling Hate Speech and Violence in India." The Washington Post. November 27, 2023. https://www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech/
23. Biswas, Soutik. 2024. "'Invisible in Our Own Country': Being Muslim in Modi's India." BBC. April 28, 2024. https://www.bbc.com/news/world-asia-india-68498675.
24. United Nations. "What is hate speech?" United Nations. n.d. https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech
25. Sightengine. "AI Image Detector. Detect AI-generated Media at Scale." n.d. Sightengine. https://sightengine.com/detect-ai-generated-images.
26. Li, Yuying, Zeyan Liu, Junyi Zhao, Liangqin Ren, Fengjun Li, Jiebo Luo, and Bo Luo. 2024. "The Adversarial AI-Art: Understanding, Generation, Detection, and Benchmarking." arXiv.Org. April 22, 2024. https://arxiv.org/abs/2404.14581.
27. Li, Yiyi, and Ying Xie. 2019. "Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement." Journal of Marketing Research 57 (1): 1–19. https://doi.org/10.1177/0022243719881113.
28. Chakrabarty, Sanjib. 2025. "Teen Beaten to Death for 'picking Mangoes'; Mob Torches Orchard, House in West Bengal." The Times of India, May 17, 2025. https://timesofindia.indiatimes.com/city/kolkata/teen-beaten-to-death-for-picking-mangoes-mob-torches-orchard-house-in-west-bengal/articleshow/121227505.cms.
29. Staff, OpIndia. 2025. "Bengal: Teenager Sudipta Pandit Killed by Sheikh Farhad for Picking Fallen Mangoes From His Orchard, Angry Villagers Burn Down Mango Warehouse of the Accused." OpIndia, May 16, 2025. https://www.opindia.com/2025/05/bengal-teenager-sudipta-pandit-killed-by-sheikh-farhad-for-picking-mangoes-from-the-ground/.
30. Voice Of Bengal. 2025. "West Bengal: Sheikh Farhad Mondal Lynched Hindu Boy to Death for Merely Taking a Mango From His Garden,." HinduPost. May 16, 2025. https://hindupost.in/crime/west-bengal-sheikh-farhad-mondal-lynched-hindu-boy-to-death-for-merely-taking-a-mango-from-his-garden-arrested/.
31. "Watch: Yogi Adityanath's Slogan 'Batenge to Katenge' | What Does It Imply?" 2025. https://www.thehindu.com/news/national/watch-yogi-adityanaths-slogan-batenge-to-katenge-what-does-it-imply/article68825689.ece.
32. Chandan Sharma on X: 'हिंदू महिलाओं को "नंगा कर मारने" की धमकी दे रहे थे, काफिरों यहाँ से भागो, ये हमारा इलाका है। शनिवार रात गाड़ी खड़ा करने को लेकर हिन्दू युवक की किसी मुस्लिम से कहासुनी हो गई थी, कुछ ही देर बाद 100-150 की तादाद में मुस्लिम भीड़ यहाँ पहुँच गई, भीड़ के हाथों में लाठी-डंडे और अन्य https://t.co/QW56cC4OUl' Dec 9, 2024. X (Formerly Twitter). https://x.com/ChandanSharmaG/status/1865982885548937717.
33. Rana, Aayushi. 2024. "Fact Check: False Claim of Muslim Crowd Targeting Dalit Hindu Families in Navsari, Gujarat." DFRAC_ORG. December 10, 2024. https://dfrac.org/en/2024/12/09/fact-check-false-claim-of-muslim-crowd-targeting-dalit-hindu-families-in-navsari-gujarat/.
34. Kaushik, Krishn. 2024. "Explainer: What Is India's Civil Code and Why Does It Anger Muslims?" Reuters, February 7, 2024. https://www.reuters.com/world/india/what-is-indias-civil-code-why-does-it-anger-muslims-2024-02-07/.
35. TNN & Agencies. 2017. "Four Wives, 40 Children, 3 Divorces Are Unacceptable, Says BJP Leader Sakshi Maharaj." The Times of India, January 7, 2017. https://timesofindia.indiatimes.com/india/four-wives-40-children-3-divorces-are-unacceptable-says-bjp-leader-sakshi-maharaj/articleshow/56390416.cms.
36. The Federal. 2023. "Fertiliser jihad, UPSC jihad: Conspiracy theories feeding hate in India." The Federal, September 29, 2023. https://thefederal.com/category/news/255-hate-speech-events-in-first-half-of-2023-80-in-bjp-ruled-states-report-97067

37. Deewan, Deepak. 2024. "एमपी में त्योहारों पर महिलाओं को फांसने की बड़ी साजिश, 14 युवक पकड़ाए तो सामने आया राज." Patrika News, September 7, 2024. https://www.patrika.com/indore-news/muslim-youths-caught-molesting-women-during-rajwada-bhajan-sandhya-at-indore-18970401.
38. Jain, Sreenivasan, Mariyam Alavi, Supriya Sharma. 2024. "The Numbers Show the Idea of Muslim Population Explosion Is Nothing but Political Propaganda." Scroll.In, May 10, 2024. https://scroll.in/article/1067694/the-numbers-show-that-the-idea-of-population-jihad-is-nothing-but-political-propaganda.
39. Rose, Steve. 2022. "A Deadly Ideology: How the 'Great Replacement Theory' Went Mainstream." The Guardian, June 8, 2022. https://www.theguardian.com/world/2022/jun/08/a-deadly-ideology-how-the-great-replacement-theory-went-mainstream.
40. Deb, Abhik. 2024. "Why Is Social Media Rife With Conspiracy Theories About 'Rail Jihad'?" Scroll.In, September 25, 2024. https://scroll.in/article/1073684/.
41. Apoorvanand. 2023. "India's Deadly Train Crash: Forget the Truth, Blame It on Muslims." Al Jazeera, June 15, 2023. https://www.aljazeera.com/opinions/2023/6/15/indias-deadly-train-crash-forget-the-truth-blame-it-on-muslims.
42. Ashok Pathak on X. 2024. "Hello @AshwiniVaishnaw जी बनाओ आप नई-नई रेल लाइन, चलाओ आप तेजस और वंदेभारत... यदि तुम निर्माण करना जानते हो तो विध्वंस करने में जिहादियों को महारत है... तोड़ने-फोड़ने, जलाने-लूटने का 1400 वर्ष का अनुभव है, इस शांतिप्रिय कौम को... https://t.co/g57KByJkzr" X (Formerly Twitter). September 10, 2024. https://x.com/pathakashokbjp/status/1833510015283696070.
43. Tiwari, Satyam and Balkrishna. 2024. "Fact Check: No 'Rail Jihad', No Vandalism. Video of Vande Bharat Repair Work VIRAL With False Claims." India Today, September 13, 2024. https://www.indiatoday.in/fact-check/story/fact-check-no-rail-jihad-no-vandalism-video-of-vande-bharat-repair-work-viral-with-false-claims-2599092-2024-09-13.
44. Maron, Jeremy. 2019. "What Led to the Genocide Against the Tutsi in Rwanda?" Canadian Museum for Human Rights, June 26, 2019. https://humanrights.ca/story/what-led-genocide-against-tutsi-rwanda CMHR
45. Das Kleine Blatt. 1939. "Cartoon Depicting Jewish Refugees as Rats, Thrown Out of Germany and Nazi-Occupied Territories, Denied Entry to Europe." The New Humanitarian, November 19, 2015. https://www.thenewhumanitarian.org/photo/20151119160019056 9/cartoon-depicting-jewish-refugees-rats-thrown-out-germany-and-nazi-occupied The New Humanitarian
46. Buerger, Cathy. 2023. "Beware of Snakes: A Common Dehumanization Trope." Dangerous Speech Project, August 25, 2023. https://www.dangerousspeech.org/libraries/beware-of-snakes-a-common-dehumanization-trope
47. Elliott, Vittoria. 2024. "'Cow Vigilantes' in India Are Attacking Muslims and Posting It on Instagram." WIRED, November 19, 2024. https://www.wired.com/story/cow-vigilantes-india-instagram/
48. Human Rights Watch. 2019. "Violent 'Cow Protection' in India: Vigilante Groups Attack Minorities." Human Rights Watch, February 19, 2019. https://www.hrw.org/report/2019/02/19/violent-cow-protection-india/vigilante-groups-attack-minorities CMHR+1
49. Saldanha, Alison. 2017. "Cow-Related Hate Crimes Peaked in 2017, 86% of Those Killed Muslim." The Wire, December 8, 2017. https://thewire.in/communalism/cow-vigilantism-violence-2017-muslims-hate-crime
50. 50.Farris, Sara R. 2017. In the Name of Women's Rights: The Rise of Femonationalism – strategic co-optation of feminist rhetoric by nationalist, xenophobic, and anti-immigration political actors. Duke University Press. https://read.dukeupress.edu/books/book/2296/In-the-Name-of-Women-s-RightsThe-Rise-of
51. The India Forum. 2022. "Flawed Logic: Karnataka Hijab Ban." The India Forum. May 23, 2022. https://www.theindiaforum.in/article/flawed-logic-karnataka-hijab-ban
52. India Hate Lab. 2024. "Jharkhand Election Brief." India Hate Lab, November 14, 2024. https://indiahatelab.com/2024/11/14/jharkhand-election-brief/
53. Mogul, Rhea, Aishwarya S. Iyer, and Sophia Saifi. 2025. "Tensions Between India and Pakistan. Here's What We Know." CNN, updated April 28, 2025. https://www.cnn.com/2025/04/24/india/pahalgam-india-pakistan-attack-explainer-intl-hnk
54. Tarun Agarwal. 2025. "Digital Aftershocks: Deepfakes in the Wake of the Pahalgam Attack in Kashmir." GNET Research, August 18, 2025. https://gnet-research.org/2025/08/18/digital-aftershocks-deepfakes-in-the-wake-of-the-pahalgam-attack-in-kashmir/

55. Jaipur Dialogues (@JaipurDialogues). 2025. "Tweet." Twitter (X), May 22, 2025, https://x.com/JaipurDialogues/status/1925599548828733855
56. Staff Author. 2025. "Civil Rights Group Documents 184 Anti-Muslim Hate Crimes in Wake of the Pahalgam Attack." Maktoob Media, May 13, 2025. https://maktoobmedia.com/india/civil-rights-group-documents-184-anti-muslim-hate-crimes-in-wake-of-pahalgam-attack/
57. Chatterji, Roma, and Deepak Mehta. 2007. Living With Violence: An Anthropology of Events and Everyday Life. London: Routledge India. https://doi.org/10.4324/9780367817640
58. Sarkar, Tanika. 2002. "Semiotics of Terror: Muslim Children and Women in Hindu Rashtra." Economic and Political Weekly 37, no. 28 (July 13–19, 2002): 2872–2876. https://www.jstor.org/stable/4412352
59. Staff Author. 2020. "The Soundtrack of Hate: Videos Glorifying Hindu Men Forcefully Putting Color on Muslim Women Shared on Social Media." Hindutva Watch, March 10, 2020. https://www.hindutvawatch.org/the-soundtrack-of-hate-videos-glorifying-hindu-men-forcefully-putting-color-on-muslim-women-shared-on-social-media/
60. Menon, Aditya. 2025. "'Zalim Hindu' Porn: How AI Is Mass Producing Pornographic Images of Muslim Women." The Quint, March 7, 2025. https://www.thequint.com/news/politics/artificial-intelligence-muslim-women-hindutva-soft-porn-images-facebook-instagram
61. Chaudhuri, Pooja. 2021. "How Twitter's Liberal Policy on Porn Allows Indian Muslim Women to Be Harassed." Scroll.in, November 5, 2021. https://scroll.in/article/1009803/how-twitters-policy-on-porn-has-been-used-to-harass-indian-muslim-women Scroll.i
62. Singha, Anand. 2025. "The Ghibli AI Trend Is Everywhere — Meet the Man Who Made It Go Viral." CNBC-TV18, March 31, 2025. https://www.cnbctv18.com/technology/the-ghibli-ai-trend-is-everywhere-meet-the-man-who-made-it-go-viral-19582046.htm
63. NPR. 2019. "Nearly 2.7 Years After Hindu Mob Destroyed a Mosque, the Scars in India Remain Deep." NPR, April 25, 2019. https://www.npr.org/2019/04/25/711412924/nearly-27-years-after-hindu-mob-destroyed-a-mosque-the-scars-in-india-remain-dee
64. Cauberghs, Olivier. 2023. "For the Lulz?: AI-Generated Subliminal Hate Is a New Challenge in the Fight Against Online Harm." GNET Research, November 13, 2023. https://gnet-research.org/2023/11/13/for-the-lulz-ai-generated-subliminal-hate-is-a-new-challenge-in-the-fight-against-online-harm/
65. Jafri, Alishan. 2025. "What the 'Cauliflower' in BJP Karnataka's X Post Means." The Wire, May 23, 2025. https://thewire.in/communalism/cauliflower-bjp-karnataka-muslim-genocide
66. Rahul, N. 2025. "CPI (Maoist) General Secretary Nambala Keshava Rao Killed in Encounter in Chhattisgarh." The Wire, May 21, 2025. https://thewire.in/security/maoist-chief-nambala-keshava-rao-killed-encounter
67. The Wire Staff. 2024. "Delhi Cop Suspended After Video of Kicking Muslims Offering Namaz Goes Viral." The Wire, March 8, 2024. https://thewire.in/communalism/delhi-cop-suspended-after-video-of-kicking-muslims-offering-namaz-goes-viral
68. Staff Author. 2025. "AI Anchors in Newsrooms: The Tech, the Business, and the Biases That Run Them." Newslaundry, March 17, 2025. https://www.newslaundry.com/2025/03/17/ai-anchors-in-newsrooms-the-tech-the-business-and-the-biases-that-run-them
69. Tiwari, Ayush. 2020. "OpIndia: Hate Speech, Vanishing Advertisers, and an Undisclosed BJP Connection." Newslaundry, June 23, 2020. https://www.newslaundry.com/2020/06/23/opindia-hate-speech-vanishing-advertisers-and-an-undisclosed-bjp-connection
70. Hinduphobia_Tr (@hinduphobia_tr). 2025. "Tweet." Twitter (X), April 24, 2025. https://x.com/hinduphobia_tr/status/1915413929796284725
71. hinduphobia_tr.2025. "Pahalgam Hindu massacre.", April 24, 2025, Instagram. https://www.instagram.com/p/DI1TaorB1gJ/
72. Hinduphobia Tracker. 2025. "Pahalgam Hindu massacre." April 24, 2025, Facebook. https://www.facebook.com/share/1J52tCUHoZ/
73. Chanchal. 2025. "Jalgaon Train Accident: What Led to the Tragedy?" LiveMint,January 22, 2025. https://www.livemint.com/news/train-accident-how-fire-rumour-killed-passengers-of-pushpak-karnataka-express-how-events-unfolded-jalgaon-all-we-know-11737552900494.html
74. Roy, Taniya. 2022. "Hindutva's Circulation of Anti-Muslim Hate Aided by Digital Platforms, Finds Report." The Wire, January 31, 2022. https://thewire.in/communalism/india-anti-muslim-hate-twitter-facebook-whatsapp-hindutva-modi-bjp

75. Mohanty, Amlan, and Shatakratu Sahu. 2024. “India’s Advance on AI Regulation.” Carnegie Endowment for International Peace, November 21, 2024. https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en&center=india
76. Staff Author. n.d. “Information Technology Guidelines Intermediaries Digital Media Ethics Code Rules 2021.” iPleaders Blog. https://blog.ipleaders.in/information-technology-guidelines-intermediaries-digital-media-ethics-code-rules-2021/
77. Mohanty, Amlan and Shatakratu Sahu, 2024. “India’s Advance on AI Regulation” Carnegie India, November 21, 2024. https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en&center=india
78. Sahaja. 2021. “Information Technology (Guidelines for Intermediaries and Digital Media Ethics Code) Rules, 2021 - iPleaders.” iPleaders. July 14, 2021. https://blog.ipleaders.in/information-technology-guidelines-intermediaries-digital-media-ethics-code-rules-2021/.
79. European Union. “Digital Services Act Article 21.” EU Digital Services Act. https://www.eu-digital-services-act.com/Digital_Services_Act_Article_21.html
80. UNESCO, 2024. “Guidelines for the Governance of Digital Platforms,” UNESCO, January 25, 2024, https://www.unesco.org/en/internet-trust/guidelines.

Understanding, Preventing, and Combating Organized Hate.

CSOH

1629 K St. NW, Suite 300
Washington, D.C. 20006
csohate.org