

Electoral Integrity in the Age of Deepfakes: What Bengal Election Reveals



Garima Sarkar & Aakriti Mittal

In the closing weeks of West Bengal's 2026 Assembly election campaign, a politically charged video spread rapidly across social media platforms. It appeared to show former Trinamool Congress (TMC) minister Humayun Kabir, then aligned with the All India Majlis-e-Ittehadul Muslimeen (AIMIM), discussing a purported ₹1,000 crore arrangement with the Bharatiya Janata Party (BJP) to split the Muslim vote. The clip spread rapidly through WhatsApp groups, Facebook pages, and X timelines, fuelling political speculation across the state. Subsequent scrutiny suggested that portions of the video had been digitally altered using AI-based editing techniques, even though parts of the underlying footage appeared to be authentic. Regardless of the precise extent of manipulation, the episode highlighted how politically charged audio-visual content can manipulate public perception long before verification or contextual clarification catches up.

The episode was not isolated. Just before the second phase of polling, another purported "sting operation" surfaced online, allegedly showing West Bengal Education Minister Bratya Basu accepting cash bribes. Shared widely, including by BJP leaders on X, the clip emerged at a strategically critical moment only hours before voting, leaving little scope for rebuttal or fact-checking before ballots were cast. The 2026 West Bengal election might ultimately be remembered not only for the BJP's sweeping victory but also for demonstrating how synthetic media and AI-generated misinformation are beginning to reshape democratic politics in India.

West Bengal has long occupied a unique place in India's electoral landscape. Since the fall of the Left Front government in 2011, the state's politics under Mamata Banerjee and the TMC have revolved significantly around coalition-building across religious and regional identities. Muslims, who constitute roughly 27 per cent of the state's population, remain electorally influential across dozens of constituencies.

For the BJP, the political challenge was therefore twofold: consolidate Hindu voters while fragmenting the Muslim electoral bloc that had helped sustain the TMC's dominance. Religious polarisation became central to the campaign. Yet the BJP's eventual victory, winning 208 out of 294 seats, cannot be attributed to digital manipulation alone.

Anti-incumbency against the Banerjee government had been building for years. Allegations of corruption, recruitment scams in the education sector, concerns regarding women's safety, and dissatisfaction over governance weakened the TMC's credibility among sections of the electorate. Since law and order, education, and public health remain state subjects within India's federal structure, these failures were increasingly associated directly with the state government. Deepfakes did not create these grievances. What they did was amplify them.

This is what makes the West Bengal election politically significant. Synthetic media functioned not as a standalone determinant of electoral outcomes, but as an accelerant within an already polarised political environment. Strategically timed videos, viral misinformation campaigns, and unverified WhatsApp forwards intensified distrust and shaped political narratives before institutional mechanisms could intervene effectively. India does have an emerging regulatory framework for addressing synthetic political content. During the West Bengal elections, the government claimed that over 11,000 social media posts and URLs were addressed through removals, clarifications, FIRs, and rebuttals under the IT Act, the IT Rules, and the Election Commission's Model Code of Conduct. The Election Commission also directed platforms to remove unlawful AI-generated content within three hours of notification.

In February 2026, the Ministry of Electronics and Information Technology introduced amendments requiring mandatory labelling and metadata traceability for fabricated content, marking India's first explicit legal framework addressing deepfakes. Yet the West Bengal election exposed a critical gap between regulation and enforcement. Most existing institutional responses, including takedowns, labels, content moderation, and post-factum clarifications, are designed for conventional misinformation campaigns. Deepfakes released immediately before polling present a fundamentally different challenge. By the time authorities identify manipulated content, the political impact may already be irreversible.

The problem becomes even more complex on encrypted platforms such as WhatsApp, where political information circulates privately and rapidly beyond the reach of conventional monitoring systems. Existing Election Commission guidelines primarily target official party accounts and public-facing content. Personal accounts, anonymous networks, and private forwards remain far harder to regulate effectively. This vulnerability is particularly concerning in a country where exposure to synthetic media

is already widespread. A 2024 McAfee survey found that nearly 75 per cent of Indian internet users had encountered deepfake content, while many identified electoral manipulation as a major concern (Verma, Sabu, and Kaul 2024). Familiarity with deepfakes has not necessarily made voters more sceptical; in many cases, it has simply accelerated circulation.

West Bengal's 2026 election might well mark the beginning of a new phase in Indian politics, one in which elections are fought not only through rallies, manifestos, and party networks, but through algorithms, manipulated videos, and viral narratives engineered to influence public perception before truth has time to respond.

The implications of the 2026 West Bengal election extend far beyond a single state contest. The widespread circulation of deepfakes and AI-generated misinformation points to a deeper erosion of democratic safeguards in India's electoral landscape, where voters are increasingly exposed to engineered information ecosystems designed to distort political reality. In the digital age, democratic legitimacy can no longer be assessed solely through electoral outcomes or the procedural integrity of the ballot box. Elections derive their true meaning from citizens' ability to make informed political choices within an informational environment that is credible, transparent, and trustworthy.

As artificial intelligence becomes more deeply embedded in political communication, future elections in India are likely to witness increasingly sophisticated forms of synthetic propaganda, targeted misinformation, and emotionally charged deepfakes aimed at influencing public perception before verification becomes possible. The danger lies not only in the spread of false information but in the gradual erosion of public trust in democratic institutions, political discourse, and the credibility of electoral processes themselves. The central challenge for democratic institutions, therefore, is no longer limited to removing harmful content after circulation, but to preventing the broader weakening of informational integrity that enables inauthentic practices to take root.

(Dr Garima Sarkar is an assistant professor of Comparative Politics and Public Policy at the Jindal School of International Affairs, O.P. Jindal Global University. Aakriti Mittal is a student researcher at the same university)