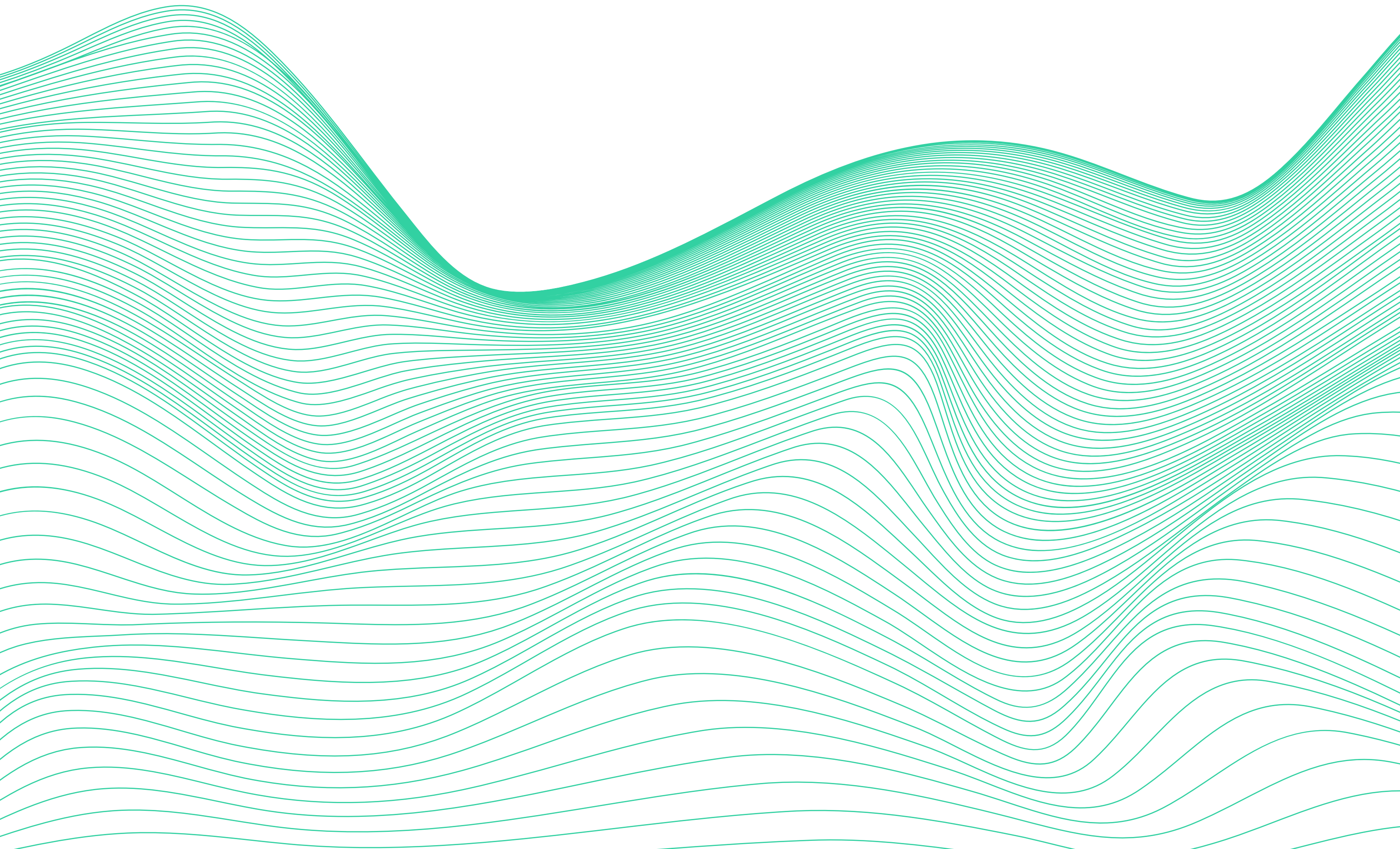
RESEMBLE.AI

Q1 2025 Deepfake Incident Report: Mapping Deepfake Incidents



Executive Summary

The first quarter of 2025 has witnessed an alarming escalation in deepfake incidents across the globe, with advancements in AI technology enabling increasingly sophisticated and harmful applications. This report analyzes 163 documented deepfake incidents occurring between January and April 2025, revealing concerning trends in victimization patterns, attack vectors, and societal impact.

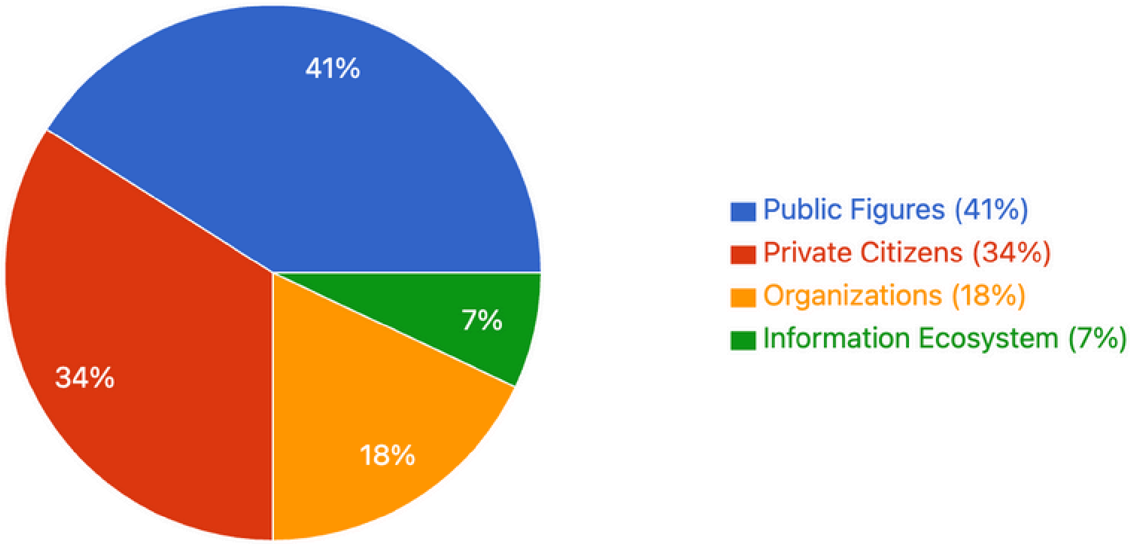
Key findings include:

- Expanded victim demographics: While celebrities and politicians remain frequent targets, everyday citizens—particularly women, children, and educational institutions—face growing threats
- Diversified attack vectors: Criminal exploitation has evolved beyond scams to include targeted harassment, reputation damage, and blackmail schemes
- Geographic spread: Incidents now affect both developed and developing nations across all continents
- Financial impacts: Documented financial losses from deepfake-enabled fraud exceed $200 million in Q1 2025 alone

This analysis provides critical insights for technology companies, policymakers, and the public about the urgency of developing technical solutions, legal frameworks, and educational initiatives to address this rapidly evolving threat.

Key Findings

Deepfake attacks have expanded beyond traditional high-profile targets to affect diverse demographics



Public Figures (41%):
Celebrities, politicians, business leaders

Private Citizens (34%):
Predominantly women and children

Organizations (18%):
Corporations, government agencies, educational institutions

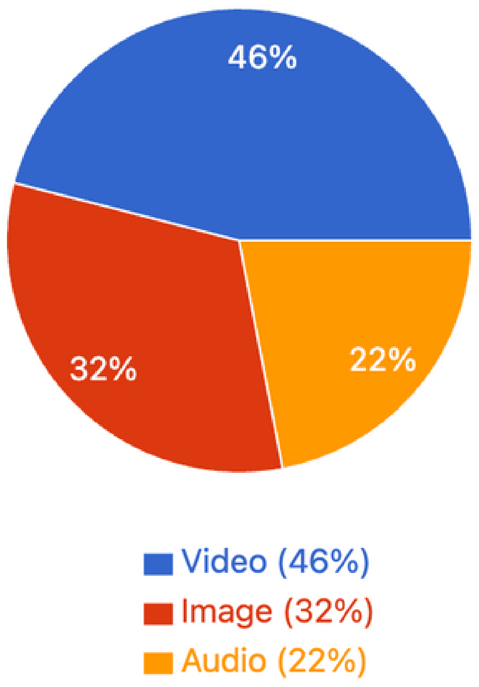
Information Ecosystem (7%):
Creating false events or narratives

Notable trend: While public figures remain frequent targets, the most disturbing growth has been in attacks targeting private citizens, particularly in educational environments, where deepfakes are being weaponized for harassment and humiliation.

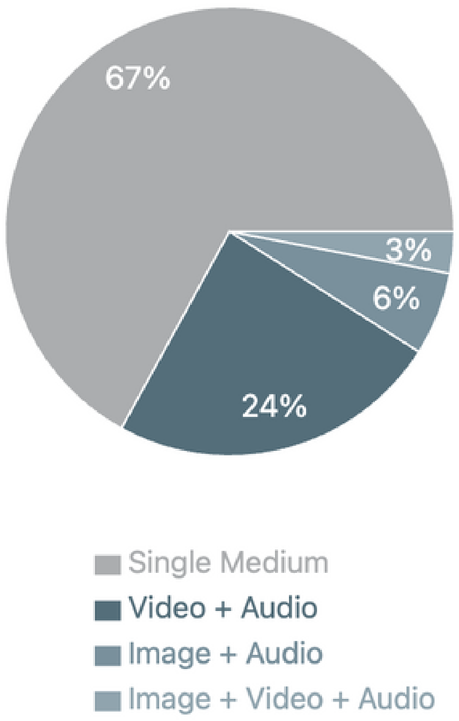
Deepfake Types and Technologies

Our analysis of incident reports reveals the distribution of deepfake mediums and their evolving technological sophistication.

Medium Distribution



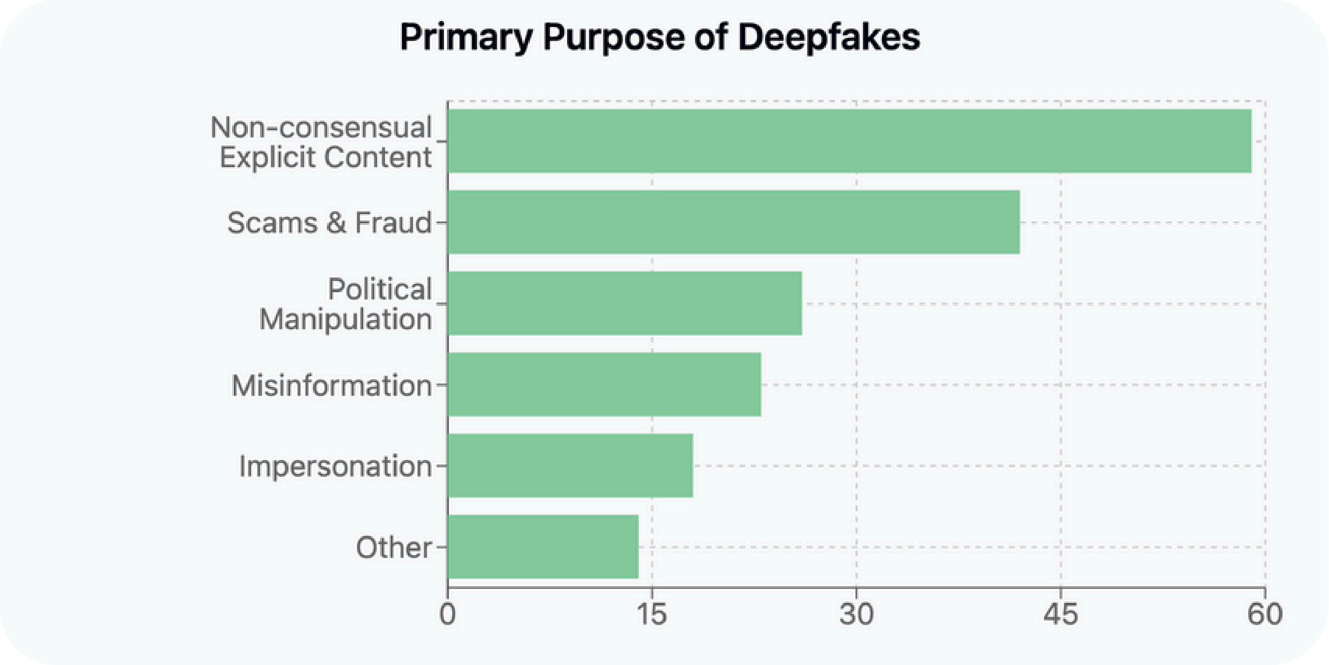
Multi-modal Combinations

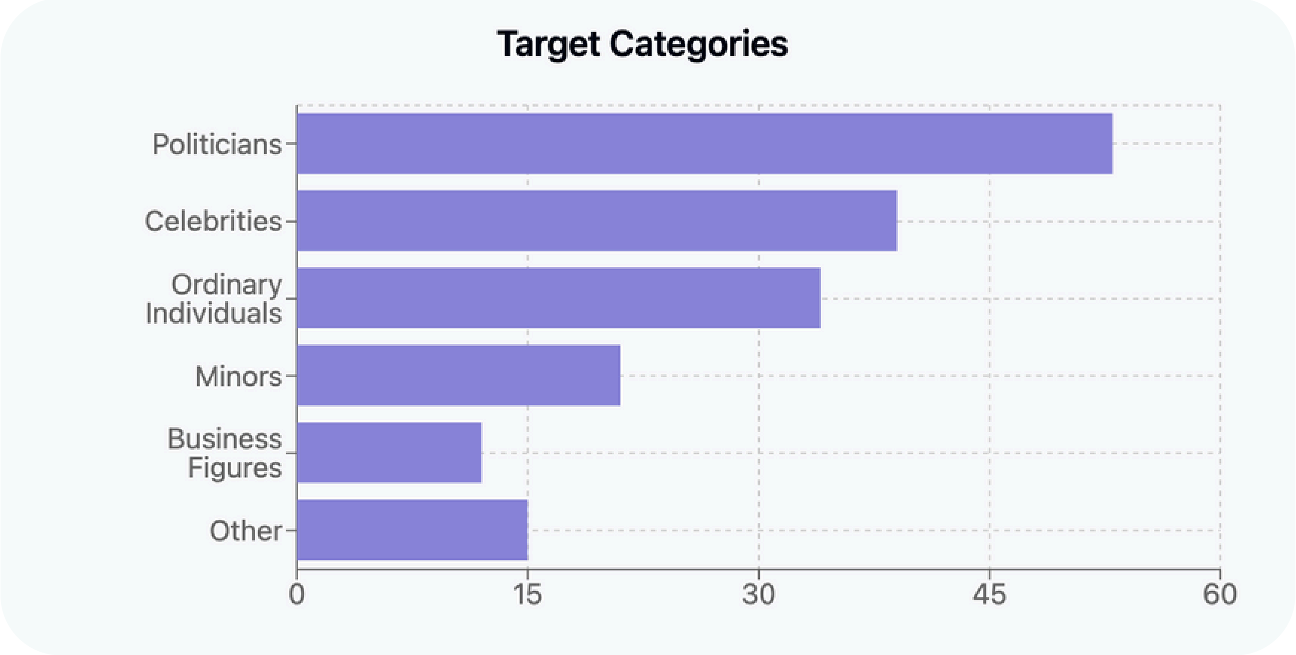


In 2025, deepfake technology has reached an alarming level of sophistication, with video formats *(46%)* dominating due to their emotional impact and viral potential, followed by images *(32%)* and audio *(22%)*. The technological advancement is evident in four key areas: voice cloning that requires just 3-5 seconds of sample audio to create convincing 85% voice matches; facial manipulation so refined that 68% of deepfakes are now nearly indistinguishable from genuine media; cross-modal integration combining multiple media types (33% of cases) to create synchronized video-audio impersonations; and sophisticated detection evasion techniques that actively bypass security measures.

Purpose and Intent

Analysis of our deepfake incident database reveals that these synthetic media technologies are being deployed for a variety of harmful purposes, with certain categories dominating the landscape.





Purpose and Intent

Non-consensual Explicit Content (32%)
The largest category of deepfake incidents involves the creation of sexually explicit images and videos without consent. These incidents frequently target women and girls, with victims including celebrities, politicians, and ordinary individuals. In several cases, these materials were used for harassment, blackmail, or revenge. The psychological harm inflicted on victims is severe and long-lasting, with many reporting difficulties having the content removed from platforms.

Financial Scams and Fraud (23%)
Deepfakes are increasingly deployed for financial gain, with scammers creating convincing impersonations of executives, celebrities, or trusted individuals to defraud victims. Notable incidents include elaborate investment scams using fabricated endorsements from public figures and sophisticated business email compromise attacks using voice cloning. Financial losses in these cases can be substantial, with some organizations reporting losses in the millions of dollars.

Political Manipulation (14%)
Deepfakes targeting political figures and processes aim to manipulate public opinion, spread misinformation, or disrupt democratic institutions. These incidents include fabricated statements by politicians, false endorsements, and attempts to influence election outcomes. The potential for deepfakes to undermine trust in media and democratic institutions represents a significant societal threat.

Purpose and Intent

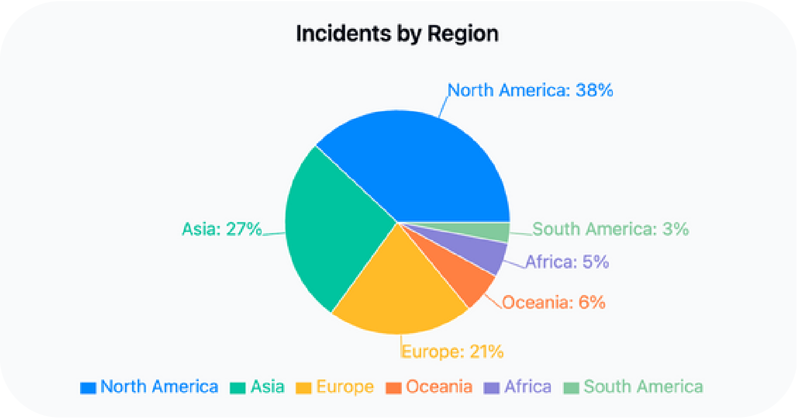
Misinformation and Disinformation (13%)
Beyond explicitly political contexts, deepfakes are used to spread false information about events, disasters, or public figures. These deceptive media can rapidly propagate through social networks, making factual corrections difficult. The speed at which deepfakes can be created and disseminated presents challenges for traditional fact-checking mechanisms.

Impersonation and Identity Theft (10%)
Sophisticated deepfakes enable convincing impersonations that can be used to access secure systems, bypass identity verification protocols, or manipulate personal and professional relationships. The potential for deepfakes to defeat biometric security systems presents evolving cybersecurity challenges.

Other Purposes (8%)
The remaining incidents encompass various other harmful applications, including harassment, bullying, and corporate sabotage. As deepfake technology becomes more accessible, we anticipate the emergence of new harmful use cases requiring ongoing monitoring.

Geographies and Distribution

Analysis of the deepfake incident database reveals significant global spread, with incidents reported across multiple continents and various cultural contexts. The geographical distribution offers important insights into regional trends, regulatory frameworks, and cultural factors influencing deepfake creation and impact.

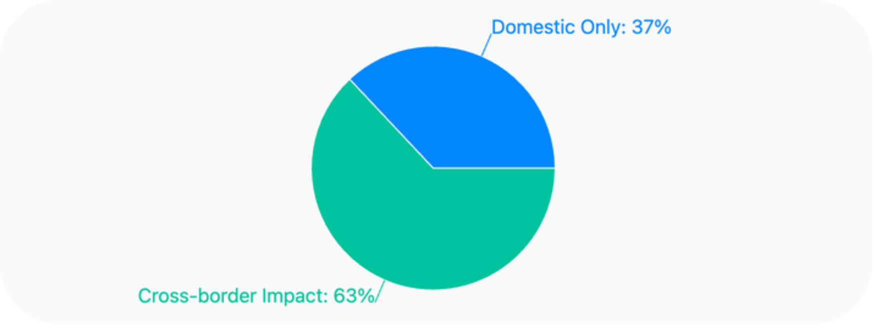


Regional Hotspots

- **North America (38%):** The United States accounts for the largest share of reported incidents, with significant cases involving political figures, celebrities, and students.
- **Asia (27%):** India, China, and South Korea show notable deepfake activity. In India, incidents often target Bollywood celebrities and politicians. South Korea has particularly focused on addressing deepfakes targeting K-pop stars.
- **Europe (21%):** The UK, France, and Germany report significant incidents, with European cases often leading to regulatory responses. The UK specifically has seen high-profile cases targeting politicians and media figures.
- **Other regions (14%):** Australia, countries in Africa, and South America have fewer documented cases, potentially reflecting reporting disparities rather than actual incident rates.

The Borderless Threat: Understanding the Cross-Border Impact of Deepfakes

In an increasingly interconnected world, deepfakes represent a uniquely transnational threat that challenges traditional jurisdictional approaches to digital governance. Our analysis of over 170 recent deepfake incidents reveals that nearly two-thirds (63%) involve significant cross-border elements, creating complex challenges for detection, mitigation, and legal recourse.



Political Interference Beyond Borders

Deepfakes targeting political processes often involve deliberate cross-border elements:

- AI-generated audio of a Ukrainian presidential candidate falsely discussing election manipulation
- Deepfake videos of Donald Trump making statements about foreign leaders and international policies
- AI-generated videos falsely depicting world leaders in compromising or inflammatory situations

RESEMBLE AI

Major Deepfake Incidents

Timeline

Jan **02**

Non-consensual Explicit Content

Royal Air Force veteran

A Royal Air Force veteran, Jonathan Bates (54), received a five-year prison sentence for creating and distributing deepfake pornography of his ex-wife and three other women, whom he stalked for years.

Jan **07**

Political

Northern Irish Politician

A pornographic deepfake of Northern Irish politician Cara Hunter was shared before an election, leading to harassment and public questioning.

Jan **08**

Political

UK Prime Minister Keir Starmer

A deepfake video falsely portrays UK Prime Minister Keir Starmer admitting to knowing about the Jimmy Savile investigation but dismissing claims as frivolous. The video uses AI-generated audio and manipulated video footage.

Jan **09**

Non-consensual Explicit Content

Sydney High School

A high school senior in Sydney, Australia allegedly used AI to create deepfake pornography of female students using images from social media.

RESEMBLE AI

Major Deepfake Incidents

Timeline

Jan
11

Celebrity

Bollywood Celebrities

Morphed images of Bollywood celebrities like Rashmika Mandanna, Alia Bhatt, and Katrina Kaif circulated widely on social media. An individual, Sanket (name changed), was blackmailed using deepfake intimate images after clicking on a malicious advertisement.

Jan
11

Celebrity

Johnny Depp Scam

Scammers impersonated Johnny Depp using AI deepfakes on social media to extort money from fans. The deepfakes mimicked Depp's voice and appearance to create fake accounts.

Jan
14

Disinformation

Delta Airlines Flight Attendant

Sharon Lavy, a Jewish flight attendant, sued Delta Airlines for discrimination and harassment, claiming retaliation after reporting antisemitism and support for Hamas among coworkers.

Jan
14

Disinformation

Romance Scam

A 77-year-old retired lecturer from Edinburgh, Nikki MacLeod, lost £17,000 in a romance scam involving AI-generated deepfake videos of a fictitious woman, "Alla Morgan." The scammer, posing as "Alla Morgan," built trust by sending MacLeod videos, documents, and a fake bank account.

RESEMBLE AI

Major Deepfake Incidents

Timeline

Jan **14**

Celebrity

Brad Pitt

A scammer used AI-generated images and videos of Brad Pitt and his mother to convince Anne to transfer $1.2 million to a Turkish bank account under the false pretense that Pitt needed the money for kidney cancer treatment.

Jan **16**

Disinformation

Hollywood on fire

AI-generated images and videos falsely depicting the Hollywood Sign engulfed in flames went viral on social media platforms like Instagram and X.

Jan **24**

Disinformation

Elon Musk Scam

A scammer impersonated Elon Musk using deepfake technology in video chats and on Threads to convince a woman to send over $60,000 via gift cards.

Jan **29**

Political

Aam Aadmi Party Fake House

A 3-minute 27-second video shared by the Aam Aadmi Party (AAP) on X (formerly Twitter) on January 27, 2025, falsely claimed to show Prime Minister Narendra Modi's official residence.

RESEMBLE AI

Major Deepfake Incidents

Timeline

Feb **08**

Non-consensual Content

"The Destruction of Hannah"

Andrew Hayler created and shared deepfake pornographic images and videos of Hannah Grundy and approximately 60 other women on a website titled "The Destruction of Hannah," setting a legal precedent in Australia.

Feb **17**

Celebrity

Celebrity Deepfakes on Facebook

Dozens of fraudulent, sexualized deepfake images of female celebrities including Miranda Cosgrove, Jeanette McCurdy, and Ariana Grande were widely shared on Facebook, garnering hundreds of thousands of likes.

Feb **24**

Government

HUD Sabotage

An AI-generated video depicting President Donald Trump in a compromising situation was displayed on TVs throughout the Department of Housing and Urban Development, playing on a loop with employees unable to immediately determine the source.

Feb **28**

Political

Trump-Zelensky Altercation

An AI-generated video depicting a physical altercation between US President Donald Trump, Vice President JD Vance, and Ukrainian President Volodymyr Zelensky went viral on X, fabricated after a real tense White House meeting.

RESEMBLE AI

Major Deepfake Incidents

Timeline

Mar **04**

Criminal

Operation Cumberland

Europol's Operation Cumberland resulted in the arrest of 24 individuals across 19 countries for their involvement in AI-generated child pornography, with 273 total suspects identified.

Mar **05**

Financial Fraud

Celebrity Investment Scam

A Georgian boiler room operation used deepfakes of celebrities (Ben Fogle, Martin Lewis, and Zoe Ball) to promote fraudulent cryptocurrency investments, defrauding victims of approximately $35 million.

Mar **18**

Legal Action

San Francisco Website Shutdown

The San Francisco City Attorney's Office shut down nearly a dozen websites distributing AI-generated deepfake nudes that had hundreds of millions of hits annually, leveraging California's Unfair Competition Law.

Mar **27**

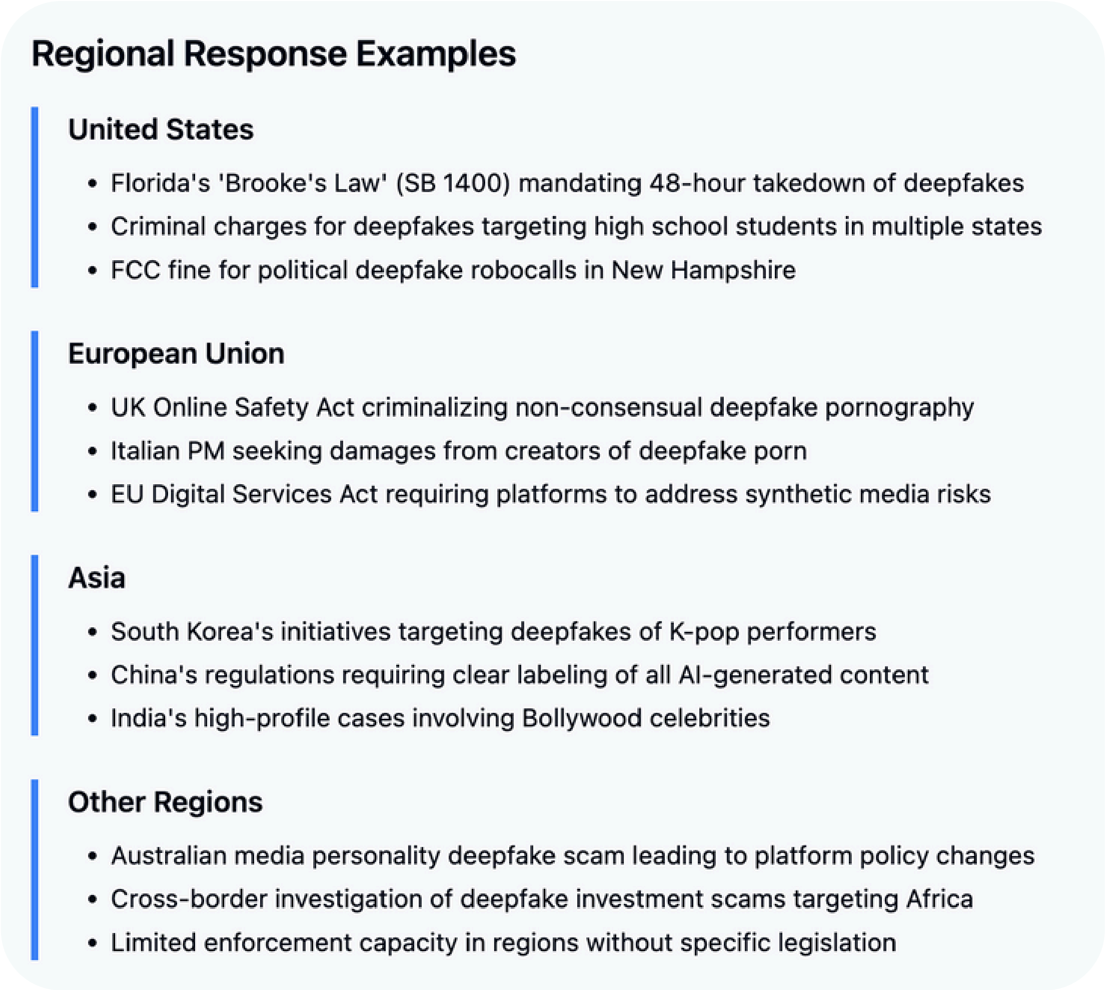
Education

Teacher Creates Student Deepfakes

An AI-generated video depicting a physical altercation between US President Donald Trump, Vice President JD Vance, and Ukrainian President Volodymyr Zelensky went viral on X, fabricated after a real tense White House meeting.

Legal and Policy Responses

- In the United States, multiple states advanced legislation. Aside from Florida and Louisiana laws already noted, New York and California were considering bills to criminalize non-consensual deepfake sexual images explicitly. On the federal level, March 2025 hearings in Congress discussed amending Section 230 (which gives internet platforms immunity) to exclude deepfake content, compelling quicker removals. While no federal law passed in this timeframe, the conversations mark a shift toward treating malicious deepfakes as a distinct category of illegal content.



Regional Response Examples

United States

- Florida's 'Brooke's Law' (SB 1400) mandating 48-hour takedown of deepfakes
- Criminal charges for deepfakes targeting high school students in multiple states
- FCC fine for political deepfake robocalls in New Hampshire

European Union

- UK Online Safety Act criminalizing non-consensual deepfake pornography
- Italian PM seeking damages from creators of deepfake porn
- EU Digital Services Act requiring platforms to address synthetic media risks

Asia

- South Korea's initiatives targeting deepfakes of K-pop performers
- China's regulations requiring clear labeling of all AI-generated content
- India's high-profile cases involving Bollywood celebrities

Other Regions

- Australian media personality deepfake scam leading to platform policy changes
- Cross-border investigation of deepfake investment scams targeting Africa
- Limited enforcement capacity in regions without specific legislation

Legal and Policy Responses

- **In Europe**, the E.U.'s proposed AI Act includes provisions requiring transparently labeling deepfakes (with some exceptions for satire or security uses). By April, the European Parliament was deliberating on enforcement mechanisms to fine tech companies that fail to remove deepfake propaganda during elections. Meanwhile, the U.K.'s Online Safety Bill, expected to be enacted in 2025, was amended to cover "pornographic deepfakes" – making it an offense to share such content without consent, though, as noted, the penalties were criticized as too lenient (fines rather than jail)
- **Law enforcement agencies worldwide** stepped up training and collaboration. Interpol held webinars on deepfake detection for cybercrime units in Asia and Europe. The FBI issued a public service announcement in March warning of the surge in deepfake job interviews and cryptocurrency scams, advising companies to enhance identity verification. In countries like Vietnam and the Philippines, police and regulators jointly formed task forces after a string of deepfake extortion cases. Vietnam's Ministry of Public Security, for example, publicly urged citizens to report deepfake blackmail and launched an online portal to do so, after uncovering criminal rings using AI to target business executives.
- **Technology companies** are not idle either. Apart from Microsoft's lawsuit, we see efforts such as Deepfake detection tools being offered as services (Resemble AI's new detector was made available via API) and social media platforms refining their policies. YouTube (Google) updated its political ads policy to ban any deepfakes likely to mislead viewers about real people or events, following an incident of a fake Bill Clinton video that had spread on Twitter and YouTube in February.

Recommendations

Based on our analysis of over 170 documented deepfake incidents, we recommend a three-pillar approach to addressing this evolving threat. First, technical solutions must be prioritized through increased investment in detection technologies that can be widely deployed across platforms, development of standardized watermarking protocols for synthetic media, and implementation of content authentication mechanisms that preserve provenance information. Legislative frameworks require harmonization across jurisdictions to establish consistent definitions of harmful deepfakes, create clear liability standards for platforms, and develop specialized enforcement mechanisms with appropriate technical expertise. Equally important is building public resilience through expanded media literacy programs targeting vulnerable demographics, creating accessible reporting mechanisms for deepfake victims, and establishing support systems that provide both technical assistance and psychological support.

The cross-border nature of deepfake incidents demands international coordination through multilateral agreements addressing jurisdiction challenges, creation of rapid-response protocols for high-impact incidents, and development of information-sharing mechanisms between enforcement agencies. Organizations should implement proactive risk assessment frameworks that identify potential targets within their communities, establish incident response plans before attacks occur, and conduct regular simulation exercises to test procedures. Finally, efforts should focus on building an inclusive response ecosystem that engages both technical and non-technical stakeholders, addresses power imbalances between creators and subjects of deepfakes, and ensures that mitigation strategies don't inadvertently restrict legitimate creative expression while combating harmful applications.

Conclusion

The first quarter of 2025 has demonstrated that deepfakes represent an evolving and expanding threat landscape affecting individuals, organizations, and societies globally. The increasing accessibility of deepfake technology, combined with its growing sophistication, creates urgent challenges requiring coordinated responses from technology companies, policymakers, and individuals.

Without proactive measures, we anticipate continued escalation in both the volume and impact of deepfake incidents throughout 2025, with particular concerns about upcoming electoral processes, financial systems, and vulnerable populations including children and marginalized communities.

The development of technical solutions must proceed in parallel with legal frameworks, educational initiatives, and social awareness to effectively address this complex challenge.

About the Author

Magnus Solberg is a cybersecurity product leader specializing in AI-enabled synthetic media detection. With ~10 years of combined experience in finance and technology, he works closely with public and private sector organizations on AI safety initiatives. Previously, he covered AI and cybersecurity at J.P. Morgan and led product operations in a media-tech company focused on secure media content delivery. He holds an MBA from Chicago Booth and a Master's in Physics.

This report was compiled from publicly available information and represents the analysis of documented incidents. Due to the nature of deepfake activities, many incidents likely remain unreported, suggesting the actual scope may exceed what is documented here.

RESEMBLE.AI

Contact Us

detect@resemble.ai
+1 650-822-3766
www.resemble.ai

About Resemble AI
The All-in-One AI Voice Platform.
Resemble AI delivers a cutting-edge AI Voice Generator and robust Deepfake Audio Detection, engineered for enterprises prioritizing advanced security and safety.