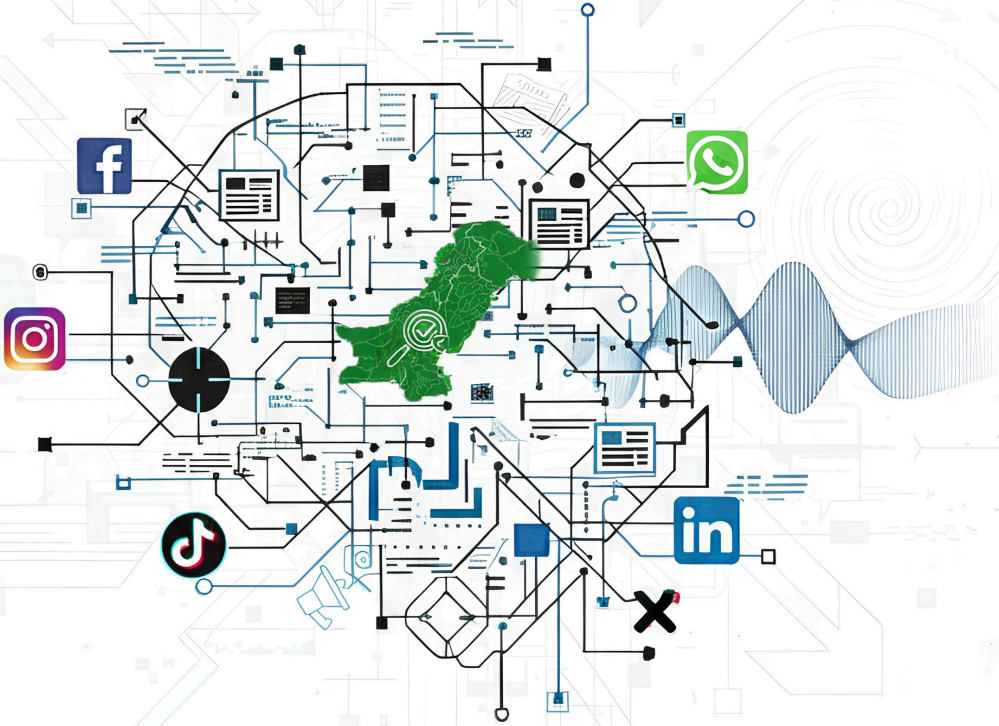
IBA CEJ
Center for Excellence in Journalism

iVerify
PAKISTAN

UNDP



AI, MISINFORMATION AND FACT-CHECKING

INSIGHTS BY iVERIFY

AI, MISINFORMATION AND FACT-CHECKING: FINDINGS AND THE FUTURE

iVerify is the United Nations Development Programme's non-partisan fact-checking project which is being implemented in Pakistan by the Centre of Excellence in Journalism (CEJ) of the Institute of Business Administration (IBA) in Karachi. It was launched in January 2024 with the mission to empower people with verified and reliable information.

2024-2025

INTRODUCTION

Checking facts is the backbone of journalism; fact-checking, however, is often overlooked despite social media and artificial intelligence (AI) rapidly reshaping information environments across the world. iVerify was launched in Pakistan ahead of the 2024 general elections as a response to the widening gap between verified and unverified information in the form of:

Misinformation: Information that is false, but not created with the intention of causing harm

Disinformation: Information that is false and deliberately created to harm a person, social group, organisation or country

Malinformation: Information that is based on real facts, but manipulated to inflict harm on a person, organisation or country

Until November 2025, iVerify monitored more than 1,000 potentially false or misleading claims – fact-checking more than 50% of them – across digital platforms and public discourse, and observed how AI in Pakistan grew to interact with a media landscape characterised by political polarisation, high social media engagement, uneven media literacy, declining public trust, and technical disruptions.

These conditions have accelerated the spread of misinformation and disinformation, making false and misleading content more viral, convincing, and difficult to detect, particularly in periods of high political activity, crises such as floods, and conflicts in the region.

Particularly during 2025, AI-enabled manipulation, systematic amplification, gendered targeting and synthetic media increasingly blurred the distinction between authentic and misleading content, underscoring the need for fact-checking as a sustained public-interest function rather than a reactive intervention.

iVerify has continued to combine editorial and public interest criteria with rigorous verification practices, monitoring and data collection to document patterns of misinformation and assess the role of emerging technologies. Our report is a testament to the project's commitment to combatting mis/disinformation and engaging with different stakeholders in doing so.

As AI is expected to play an even more active role in 2026, the findings presented here (detailed in our report) highlight the importance of strengthening fact-checking through strategic partnerships with media organisations, academic institutions, and development actors. Sustained collaborations will be critical to building resilient, credible, and accountable information ecosystems in Pakistan and across South Asia.

FIND THE FULL REPORT WITH INSIGHTS, DATA, ANALYSIS AND RECOMMENDATIONS ON

IVERIFYPAKISTAN.COM

THE FACT CHECK TEAM

ZAHRAH MAZHAR
PROJECT LEAD

SYED TALAL AHSAN
PROJECT SUPERVISOR

AINUM BAIG
SENIOR FACT CHECKER

MUZHIRA AMIN
FACT CHECKER

MUNEEBA IQBAL
FACT CHECKER

MASOOD
FACT CHECKER

ROSHAIL KHAN
SOCIAL MEDIA MANAGER

SAROJ RIZWAN KHAN
PROJECT COORDINATOR

METHODOLOGY

iVerify Pakistan follows a four-step methodology for fact-checking that begins with monitoring and ends with publishing an evidence-based fact-check.

How do we select claims: The claims found in monitoring are evaluated for either further investigation or are discarded.

Criteria for claims pursued: Virality, public interest, potential harm and relevance to the ongoing news cycle. Some claims, depending on their significance in the context of news, public interest and projection of their circulation based on similar past claims, might be selected for pre-emptive debunking, despite low virality, to mitigate and control the potential damage they might cause.

Criteria for claims discarded: Lack of access to a place, documents or response from authorities; limitation in investigative resources; inconclusive findings to avoid adding to the confusion; evaluation of 'more harm vs good' by sharing a piece of unverified information – even in the form of a fact-check – when the claim can pose danger to someone's life, expose personal details, etc.

What is the two-step verification? The selected leads, based on their nature, are investigated using visual verification and forensic tools, corroborated with official sources and consulted with reporters and experts. The two-step verification means that after the fact-checker assigned to the claim has completed their investigation, another fact-checker as well as the Project Supervisor/Project Lead reviews the process to ensure accuracy before publication.

iVERIFY'S FACT-CHECKING PROCESS

1. MONITOR

- Social media
- TV
- Print
- News websites
- Tips from public
- Public figures

2. ANALYSIS

- Virality
- Harm factor/ Hate speech
- Public interest

3. INVESTIGATION

- Reverse image search
- Keyword search
- Media corroboration
- Forensic & AI analysis
- Expert consultations

4. PUBLISHING

- Two-person verification
- Media partners/curation
- iVerify website
- iVerify social media platforms

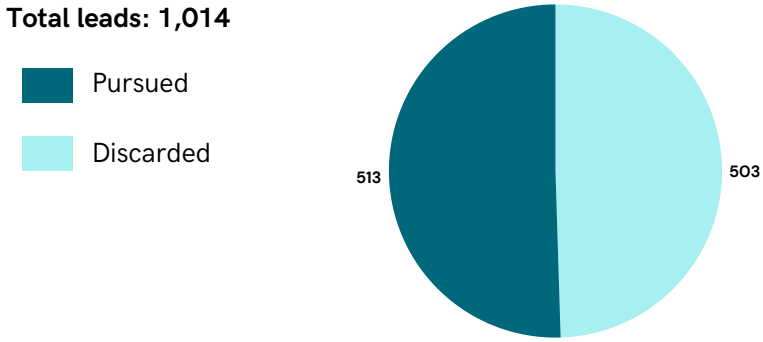
iVERIFY TOOLKIT

Tools play a key role in iVerify's fact-checking process but bearing in mind that they can also give false/inconclusive results, the team keeps using different tools and relies on more than one to provide as evidence. Particularly for AI-generated content, tools are used as supportive aids that give likely indications about the status of a piece of content, rather than any conclusive results, and have to be used in conjunction with traditional visual and logical analysis.

NAME OF TOOLS	IMAGE	VIDEO	AUDIO	AI	LOCATION	ARCHIVING
Google Lens	✓	✓				
Hive Moderation				✓		
Sight Engine				✓		
Attestive.ai				✓		
AIorNot				✓		
InVid	✓	✓				
FotoForensics	✓					
Fotoforensically	✓					
Decopy.ai				✓		
Wasitai.com				✓		
Undetectable.ai				✓		
Fake image detector.ai				✓		
Fake image detector.com	✓					
illiminatory.ai				✓		
Google Maps					✓	
Google Street View					✓	
Deepfake Total			✓			
Perma						✓
Wayback Machine						✓
Archive.ph/md						✓
Ghost Archive						✓
TruthScan				✓		
Deepware				✓		

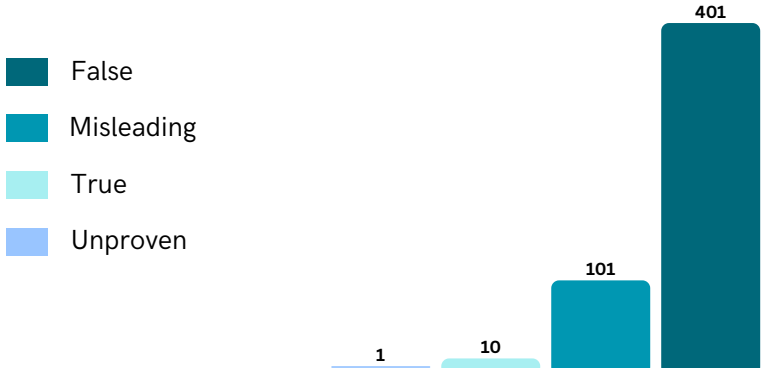
1. NUMBER OF CLAIMS PURSUED AND DISCARDED

Between December 2023 and November 2025*, iVerify Pakistan identified a total of 1,014 leads, of which 513 were pursued, and 501 were discarded.



2. BREAKDOWN OF PURSUED CLAIMS

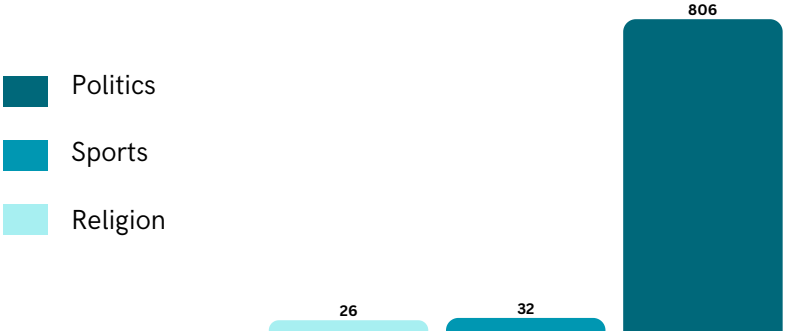
The majority of the fact-checks carried out were found to be false, followed by those found to be misleading.



*Data for all graphs is from this duration

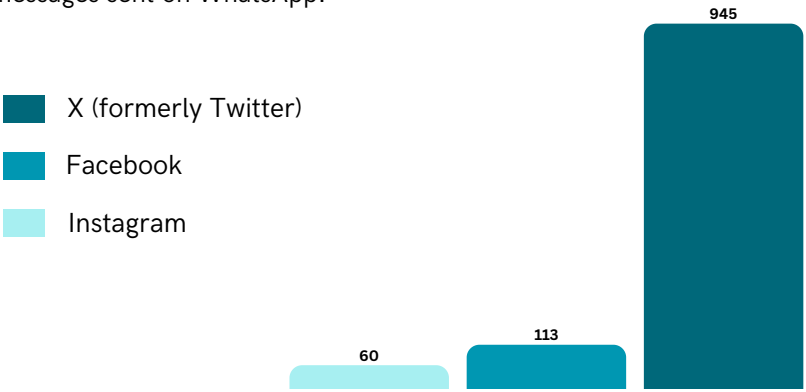
3. TOP THREE TOPICS

The top three topics that dominated the claims identified by iVerify Pakistan were politics (806), sports (32), and religious (26).



4. TOP THREE SOCIAL MEDIA PLATFORMS USED TO CIRCULATE CLAIMS

The top three platforms most frequently used to spread mis-disinformation identified by iVerify Pakistan were X (945 claims), Facebook (113 claims), and Instagram (60 claims). It is pertinent to mention that data cannot be collected on messages sent on WhatsApp.

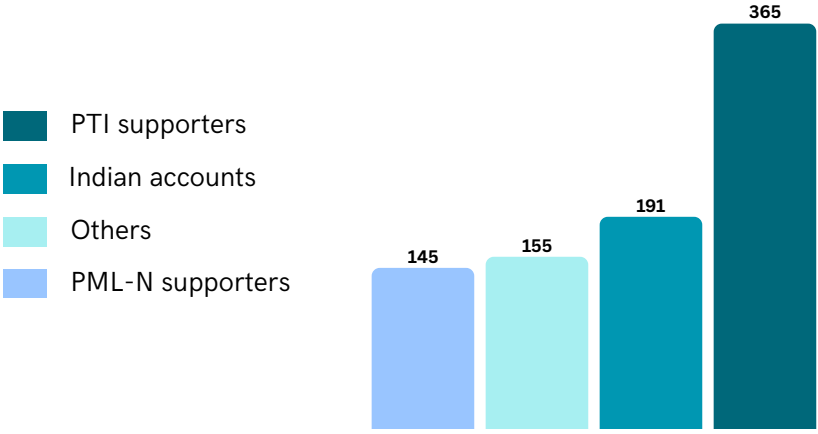


5. TOP FOUR PERPETRATORS

Metrics for 'Perpetrators' of the claims identified by iVerify are tricky as they can't take into account statements by government bodies or politicians that are shielded by "state security" or the element of confidentiality; similarly, news stories or journalists quoting unnamed sources pose the same challenge for fact-checkers.

Our data, based on the leads found, show PTI supporters (365 claims) to be a close second to the Others (x) category – which includes media, politicians, government officials and institutions, celebrities, religious supporters and more – making them the most active group in this sphere. The other two groups which were noticeable in spreading mis/disinformation were Indian accounts (191 claims) and PML-N supporters(145 claims).

Supporters of political parties were determined by considering factors such as their past posting habits and visual information from their profiles, including their names, handles and display images. Indian accounts were a conglomerate of propaganda accounts and users, those sharing pro-Indian and anti-Pakistan material with the location on X set to India, right-wing accounts and journalists.

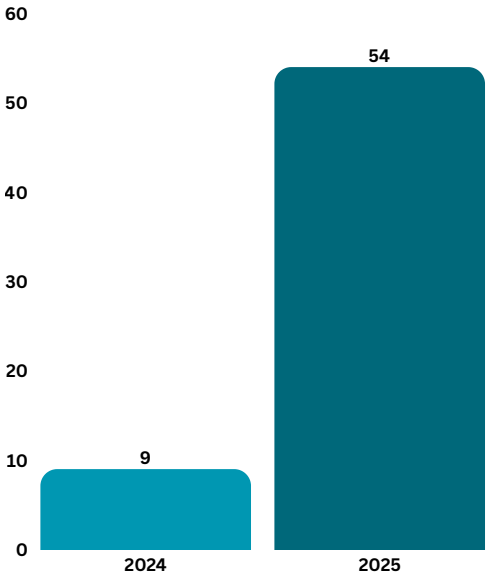


ARTIFICIAL INTELLIGENCE AND MIS/DISINFORMATION

The capabilities of AI are both a bane and a boon. For all the creative possibilities that AI unlocks, it has become a tool for misinformation and disinformation on both domestic and global levels.

iVerify's interaction with AI-related mis/disinformation started slow; there were relatively low numbers of leads in 2024 while 2025 saw a spike, mirroring real-world developments in this technology. Not only was AI used more widely, including as part of policy in organisations, but the rollout of more intuitive, user-friendly and advanced platforms, easily and freely available to the public, allowed more people to create realistic visuals.

GROWTH IN AI-RELATED FACT CHECKS (2024 VS 2025)



From nine fact-checks related to AI in 2024, iVerify witnessed a jump to 54 in 2025, demonstrating the increasing adoption of the technology in the mis/disinformation landscape, particularly as it supplants older and cruder forms of image and video manipulation.

Furthermore, from initially being used to target politicians, particularly females, iVerify saw how AI-related mis/disinformation soon encompassed various domains, including sports, entertainment, climate and business.

With more skilled AI-generated/assisted claims emerging in 2025, iVerify Pakistan has had to constantly reevaluate fact-checking mechanisms and update them as advances in the technology are rolled out.

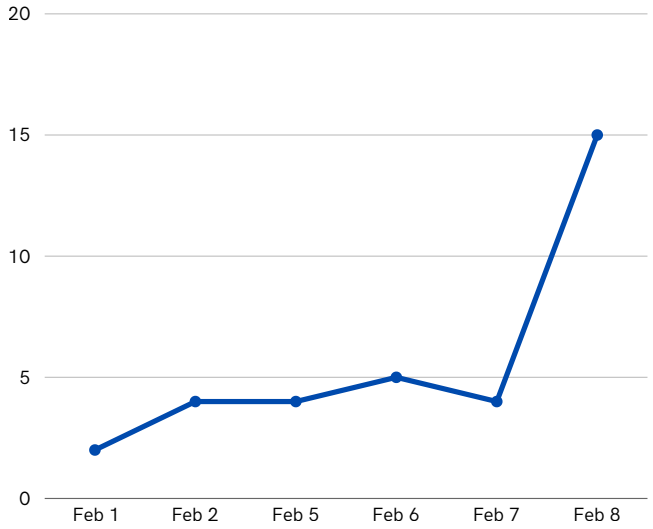
A dangerous trend spotted by iVerify is the "AI-fication of reality": when AI is used to make changes to otherwise original visuals, sometimes in the slightest but enough to mislead the public. Correspondingly, this has led to a growing awareness in the public about the spread of AI content, with people increasingly asking chatbots to verify certain visuals or calling out deepfake content.

USE OF AI IN PAKISTAN GENERAL ELECTIONS 2024

Interestingly, the advanced use of AI among the claims disseminating mis/disinformation leading up to the 2024 elections remained low, especially when compared to other countries undergoing elections that year.

It's noteworthy that during the campaign period, the PTI, leveraging its strong social media presence, strategically incorporated AI tools, as most of its prominent leadership was behind bars and the party was facing multiple hurdles. Allegations even surfaced suggesting PTI founder Imran Khan's piece for *The Economist* might have been AI-generated.

STATISTICS OF CLAIMS IN ELECTION WEEK (FEBRUARY 2024)



A sharp spike was seen on election day, despite internet disruptions, showing the high engagement and interest of public.

iVerify found some claims which involved images that had been tampered with, but none with a highly technical use of AI. On January 30, less than two weeks before the elections, when Khan and his close aide Shah Mahmood Qureshi were convicted in the cipher case, a manipulated image of the PTI founder sporting a beard began circulating online.

While the public's interest in PTI founder's appearance following his incarceration was high, online discourse showed that scepticism around the photo was also high. Investigation confirmed that the image was indeed doctored, with Fake Image Detector saying it was either computer-generated or modified.

The most dangerous trend spotted was the intention to interfere in the electoral process by confusing or misguiding voters. Case in point: a purported audio of Khan announcing a boycott of the elections circulating online. It was, however, immediately debunked by the PTI. Similar attempts, although not necessarily using AI, were made via religiously charged posts against political opponents, false claims, frequent recurrence of misinformation regarding the withdrawal of candidates from the electoral race, and other misleading claims, as explored further in the report.

AI IN POLITICAL TARGETING

Pakistan is a politically charged country, and politics dominate most public discussions; leads related to AI were not immune to this trend either. iVerify observed that AI-generated content and visuals on political developments and politicians were primarily spread by party supporters to bolster and prop up their leaders while spreading propaganda to undermine and diminish opponents.

The common players were supporters of the PTI and the PML-N. In the PTI's case, the party's campaigners used key international events such as the US elections to build on their narratives using AI, i.e. the dissemination of doctored videos of US President Donald Trump and his aide Richard Grenell. The videos were used to lend the impression that they were voicing support for Khan. Conversely, for the other side of the camp — the ruling PML-N — Khan was a key target. iVerify Pakistan noted multiple instances of the ex-premier's old videos being manipulated to push an anti-PTI narrative.

Across both sides of PTI and PML-N supporters, the use of AI-generated visuals to spread scandalous content about female politicians was another common observation, albeit the crude nature of such content lends to low credibility with people seen calling such content out. However, the nature of such claims makes them dangerous even with low believability.

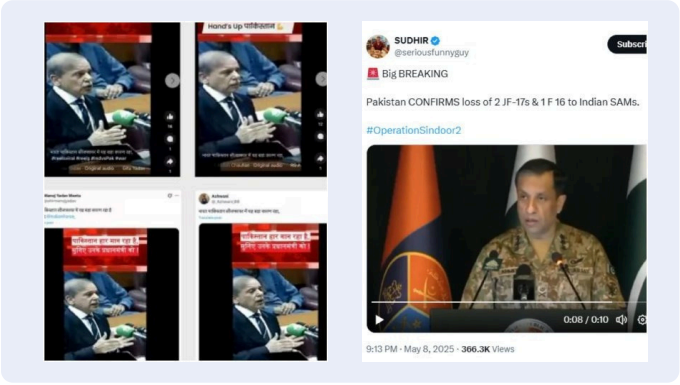
But beyond rivalries, iVerify observed another trend wherein misinformation on political events was spread using exaggerated AI imagery. At the same time, generative technology was also used to push a particular narrative by dubbing videos of government representatives in key positions.

In March 2023, PM Shehbaz became a target of this trend when a viral video allegedly showed him being confronted over funding Umrah for a 30-member government delegation. When investigated, it was found that the clip was dubbed over.

WARS, CONFLICTS AND AI

In today's world, wars are no longer just fought on battlefields; they have now moved to the reel world, where truth has become both a casualty and a weapon. During its operations between 2024 and 2025, iVerify closely monitored how AI is now being weaponised.

In this regard, the 2025 Pakistan-India conflict particularly demonstrated the use of AI-generated content in shaping public perception. A striking example was a deepfake video of Prime Minister Shehbaz Sharif, which was edited to show him conceding defeat to India. Another such example was an altered video of Pakistan's military spokesperson allegedly claiming the loss of two Pakistan Air Force fighter jets.



Manipulated clips that were used to spread misinformation regarding statements issued by the prime minister and the military spokesperson gained millions of views and were reshared hundreds of times.

In these instances, and many more, iVerify noted how AI-generated content was used to deliver a narrative that is entirely fabricated. Especially in times of conflict and war, when information is dumped on the internet at an overwhelming speed and quantity, it becomes easier for users to believe in any and all multimedia posted online, as the line between fact and fiction blurs.

At the same time, there is a lack of access in these situations for the media, so the public can only consume official versions from both sides; this leads to higher believability in what is being spread through public discourse, as there is a massive appetite for information at a time of war and uncertainty. In the case of Pakistan and India, people from both sides were on the lookout for what each country was saying – beyond official statements – adding to the virality of unverified information. Closer coordination between the government and fact-checkers could assist in this misinformation whirlpool.

A similar but more intense trend was once again seen during the Iran-Israel conflict of 2025. iVerify Pakistan traced eight such leads during the 12-day war, where generative tools were used to create seemingly authentic content in real-time conflict zones. One widely shared post featured an AI-generated image of mid-rise buildings collapsing in Israel, claimed to be the result of Iranian missile strikes. In reality, though, the clip was created using synthetic imagery.

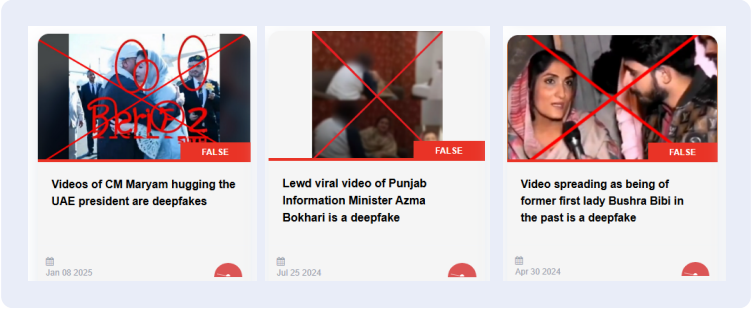
Videos showing Iranian forces thanking Pakistan for intelligence support were also shared extensively on Facebook and X. However, these videos were entirely AI-generated, misleading audiences about Pakistan's role in the conflict.

AI AGAINST WOMEN

Pakistan is regularly placed at the bottom of international indices related to women's rights, appearing at 148 out of 148 countries in the Global Gender Gap Report in 2025. Enter a host of powerful and unsupervised AI tools, functional on existing gender biases, now enabling that violence to proliferate in complex ways. During its two-year operations, iVerify identified a trend where AI was used to create new forms of abuse and amplify existing ones, using women as softer targets, especially those under the public eye.

The purpose is clear: to undermine their credibility, a tactic that is increasingly being used by political players against women of the rival camps. The way it is done follows a well-thought-out process, which involves using AI to generate scandalous videos and photos. These clips are then disseminated from not just a few but dozens of accounts across social media, each gaining views in hundreds, thousands, even millions at times.

Punjab Minister for Information and Culture Azma Bokhari became a victim of this new form of abuse in 2024, when a sexualised deepfake video began circulating. And despite being in an important position in a key province, the online harassment and character assassination that came her way did not fall short.



A collage of fact-check reports on AI-generated deepfakes that targeted female figures of opposing political parties.

Punjab Chief Minister Maryam Nawaz and former first lady Bushra Bibi are the other two common targets, whose fake videos and images pop up on social media every time the political parties affiliated with them lock horns.

This online harassment is not just restricted to the political sphere; in 2025, journalist Benazir Shah was targeted when a deepfake video claimed to show her dancing inside a nightclub. Earlier, iVerify noted a trend of manipulated videos showing mainstream Pakistani actresses in intimate scenes — all of this content was entirely fake.

These campaigns not only inflict emotional distress and reputational harm upon the targeted individuals but also perpetuate a culture of toxicity and misogyny within public discourse. It serves to intimidate and silence voices, particularly those of women, thereby impeding their ability to participate fully in public life and contribute to meaningful dialogue.

AI AND 2026: WHAT'S AHEAD

"We're going to move from assuming what we see is real by default, to starting with scepticism. Paying attention to who is sharing something and why."

This insight was among Adam Mosseri's, the head of Instagram, reflections on the last day of 2025. iVerify observed this trend increasingly in the second half of the year; as the use of AI became more and more rampant, a parallel rising trajectory – while not at the same speed – of the public paying attention to the content and flagging it as "fake" also became more and more apparent. At the very least, people are questioning, more so than before, whether a visual is authentic or fabricated. Information, however, is still being consumed in the same manner, which means that media outlets have a bigger role to play by claiming more space on social media. There is a possibility that fact-checking will pivot towards flagging authentic content more than verifying false claims. Misleading information and visuals will require the most attention, as they have the potential to deceive even those looking for evidence of AI-tampering.

This is all to say that the future of AI and its impact on mis/disinformation is anything but predictable or straightforward.

Experts highlight that AI-generated deepfakes, voice cloning, and synthetic images are approaching near-indistinguishability from authentic content. This raises the question whether social media platforms will monitor and regulate AI labels aggressively – and how. Will governments and legislation take the matter of transparency up?

"AI trust paradox" is a phenomenon global researchers have already identified; as AI outputs become more persuasive and audiences struggle to differentiate fact from fabrication, trust in both news media and digital platforms will erode further. This is where the possibility emerges that trusted media and platforms will be sought out by the public once again, with the need to know what is real amidst the inundation by AI slop outweighing the trust deficit in mainstream media.

HATE SPEECH

iVerify, throughout its operations, has noted harmful content's constant presence on the internet, manifested in different ways, including gender, ethnicity, religion, and political affiliation.

While harmful content is broadly damaging for those it targets, content designated as hate speech goes a step ahead in the vitriol it expresses and the danger it holds.

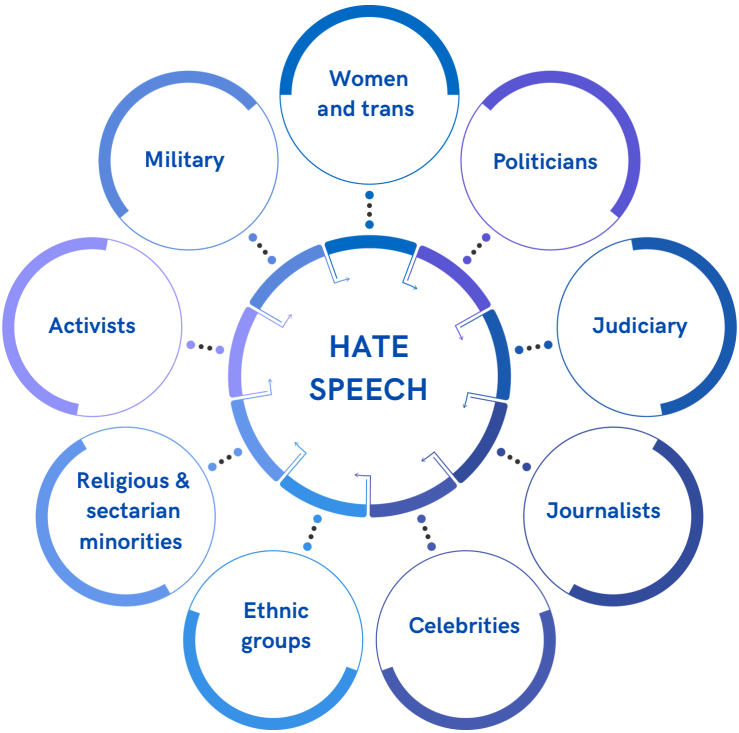
The United Nations describes hate speech as "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor".

A core function of iVerify since its launch has been to monitor such content due to its extreme nature; the team observed that harmful content and hate speech directed at political parties and figures often took on a religious tone as those spreading the incendiary material capitalised on the religiously charged sentiments of society. For instance, posts perpetuating links between the Ahmadi community – referred to by the derogatory term 'Qadiani' – and PTI leaders were fact-checked and observed to spread only for the purpose of inciting hate.

Harmful content and hate speech were a mainstay in claims pertaining to the judiciary, particularly during times of politically charged cases in 2024. The team observed sustained campaigns against the courts, particularly against former chief justice of Pakistan Qazi Faez Isa or those judges considered to be biased against Khan by PTI supporters.

This campaign and environment of negativity towards the Supreme Court came to a head in February 2024 when JUI-F chief Fazlur Rehman incorrectly accused Justice Isa of "allowing the Ahmadi community to spread its beliefs" after a court verdict.

This tactic of vilifying individuals or institutions based on religious or ideological differences is not only divisive but also undermines democratic principles and pluralism. It is important to note that hate speech has been observed to target multiple groups of people in Pakistan's society (details can be found in our report online).



CHALLENGES OF FACT-CHECKING

Fact-checking as an enterprise faces many of the same challenges that the wider media landscape faces in Pakistan: a lack of cooperation from state officials and institutions on certain topics, such as those related to security and the military, being entirely off the table, to the impact of internet and social media clampdowns on investigative efforts.

Among the more particular challenges to fact-checking itself, a major one pertains to timing and delivery. Spreading a credulously fabricated visual with an accompanying caption is a matter of a few minutes, while the process of investigating it may be a longer enterprise.

At the time, that content might have spread to various platforms and already achieved the desired effect, while the corresponding fact-check report that debunks it is published online with a lag and a more limited delivery mechanism.

Related to the above, another challenge deals with the recurrence of mis/disinformation, whereby the publication of a fact-check report does not itself put an end to the spread of a piece of content. It has been observed that certain perpetrators spread the same old images and videos with the same claims repeatedly, many times over a period of time, while accruing high views. Fact-checking a piece of content, therefore, not only requires investigating it once but also regular monitoring to preemptively respond to it when it is shared next.

Another major challenge that dominates the focus of fact-checking efforts is related to the investigation and analysis of AI-generated content. Earlier attempts at such visuals had certain telltale signs or errors that made it easier to spot them and point them out to audiences. However, with the march of technology and increasing sophistication of newer programmes, AI-generated content is fast approaching a level of detail and quality that overcomes many of these earlier errors.

Compounding the issue is that the reliability of tools to flag and detect AI-generated images and videos lags, with no single multipurpose tool still that provides a high degree of reliability for all cases.

Investigating through the use of forensic tools thus often requires testing on multiple tools, with the primary method being to still look for visible flaws in a piece of content or attempt to dismantle its logic.

LOOKING AHEAD: IVERIFY IN 2026

In 2026, iVerify aims to play a pivotal role in putting higher value on accurate content and mitigating the risks of AI's sophisticated use and high social media penetration by forming strategic partnerships with

- Journalism fraternity
- Academic institutions
- Media development groups

Looking ahead, the correlation between AI-driven misinformation and media integrity underscores the importance of exploring hybrid human-AI fact-checking frameworks, as experts have pointed out, leveraging automated detection, collaborative research, and public engagement. Fact-checkers, assisted by the combination of AI detection tools, editorial skills, and media literacy initiatives, will be essential in building resilient information ecosystems that can preemptively counter misinformation.

iVerify's work is designed to serve multiple stakeholders and purposes:

- newsrooms seeking verification support
- researchers analysing patterns of mis/disinformation
- private and government bodies working to strengthen democratic governance and social cohesion
- strengthening fact-checking skills of media students and trainers
- design data-driven and evidence-based media literacy campaigns for the public
- training journalists and other professionals in fact-checking and countering false information effectively
- community awareness and engagement through programmes like iVerify Advocates and iVerify Tipline
- collaborations with international fact-checkers, including DW, AFP, and Alt News, as a local reporting body

As AI becomes a more active and influential force by 2026, the sustainability of fact-checking will depend on deeper collaboration between media organisations, academic institutions, and the development sector. This publication is both a reflection of what has been learned so far and an invitation to partners committed to safeguarding credible, inclusive, and accountable information environments in Pakistan and beyond.

'I VERIFIED': JOIN US

Students: become iVerify advocates

Academic institutions: enhance fact-checking skills for media students and teachers through trainings by iVerify; receive support on incorporating fact-checking as a subject into curriculum

Media: collaborate on fact-check investigations and social media campaigns via the dissemination of material, ensure upgradation of journalists from your newsroom through fact-check trainings and skill development, send employees to participate in fellowship programmes to become iVerify team members

Local and international fact-check platforms: open up collaboration to share use of resources and assets, pursue joint investigations best utilising respective strengths for greater depth and breadth, exchange employees to learn best practices

Government bodies: fact-check training and skill development;
closer collaboration to contain the spread of mis/disinformation

Corporate sector: run awareness campaigns on the importance of verified information in the age of AI as part of CSR initiatives

Public: share verified content in close-knit circles and communities and be an active member of the iVerify initiative by contributing claims, tips, and stories that you come across.

If you wish to verify any news, image, video, or claim and check whether it is true or misleading, visit our website and navigate to "Submit a Story." Please include the relevant link along with brief context.

SUBMIT YOUR STORY/TIPS ON OUR WEBSITE
IVERIFYPAKISTAN.COM

OR

DROP US AN EMAIL AT
IVERIFYPAKISTANATCEJ@GMAIL.COM

You can also stay updated by following us on our social media platforms, where we share verified information, explainers, and updates.

@IVERIFYPAKISTAN

SOCH PHIR POST

SCAN TO STAY CONNECTED

iverifypakistan.com

Front cover was made using AI