Universidad Externado de Colombia Law school

BLOG DE DERECHO DE LOS NEGOCIOS

Business Law.

August 1, 2025

Virtual hearings at risk: deepfake technology and the breach of judicial integrity

By: Nicolas Ernesto Lozada Pimiento and Ed Segura

Nicolás Ernesto Lozada Pimiento[1]and Ed Segura[2]

Summary

The rapid integration of videoconferencing technology into court proceedings, particularly accelerated by the COVID-19 pandemic, has brought significant efficiency and accessibility to judicial systems. However, the rise of deepfake technology introduces substantial risks that threaten the integrity of court proceedings. This paper examines the vulnerabilities introduced by deepfake technologies, assesses their potential impact on judicial fairness and reliability, and recommends comprehensive protective measures to address these emerging threats. It also argues that physical presence in the courtroom is vital to maintaining judicial authenticity and credibility.

1. Introduction

The global judiciary has undergone a significant transformation with the widespread adoption of virtual proceedings facilitated by platforms such as Zoom, Webex, and Microsoft Teams. Initially a temporary solution during COVID-19, these virtual hearings have become permanent due to their practicality and cost-effectiveness (National Center for State Courts, 2021). However, this technological transition has exposed the judiciary to new threats, notably deepfake technology, an advanced synthetic medium powered

by artificial intelligence that convincingly mimics real people, with serious implications for judicial integrity.

2. Understanding deepfakes

Deepfake technologies use generative adversarial networks (GANs) to generate hyper-realistic synthetic audio and video. GANs consist of two neural networks: the generator creates synthetic media, while the discriminator assesses authenticity. This iterative process results in content that is nearly indistinguishable from reality (Deeptrace Labs, 2020).

[Figure 1: Growth of deepfake content (2019-2023)]

Deepfake media is found primarily in images and videos, chats, and audio content.

Regarding images, "face swapping" is used when a person's face is replaced with another image, simulating that the person seen is different from the real one. On the other hand, deepfake videos can be created using three techniques: lip-syncing, facial synthesis, and attribute manipulation (Nguyen et al. 2019b; Masood et al. 2023).

The second type of deepfake is text-based deepfake, used primarily on social media for fake comments and reviews on e-commerce websites.

The third type of deepfake is known as audio deepfake. These deepfakes involve the use of AI to create a synthetic, lifelike human voice. These deepfakes can be created using text-to-speech or voice-swapping methods.[3]

Deepfakes have already been used in political disinformation, fraud, and identity theft. Prominent examples include manipulated speeches by political figures designed to deceive voters (Microsoft, 2021) or, more recently, Google's artificial intelligence-powered video creator Veo 3 2025, which makes it difficult to identify whether something is real or not.[4]

3. Application of videoconferencing in judicial systems

Videoconferencing technology is widely used in court settings for remote testimony, virtual appearances, and international cooperation. Courts have benefited considerably by increasing the participation of geographically distant or vulnerable witnesses (National Center for State Courts, 2021). In arbitration, a form of alternative dispute resolution, in 2019, theInternational Centre for Settlement of Investment Disputes (ICSID)reported that around 60% of its hearings and sessions were held remotely (Simson, 2020).

Prominent examples include the increased rates of witness testimony in family and criminal courts. In recent years, videoconferencing has also played a crucial role in international arbitration hearings, especially in cases where saving time and costs is a priority for the parties involved.[5]

However, inadequate biometric verification and authentication protocols, as well as limited public investment in technological systems, make virtual courtrooms susceptible to exploitation through deepfakes, such as identity fraud and evidence tampering. Improved biometric identification methods, such as multifactor authentication and live verification, are urgently needed to safeguard judicial processes.

4. Deepfake threats in court hearings

Deepfakes pose a critical threat to legal proceedings, in particular:

- **Impersonation:** A high-profile child custody case in India involved an individual who used deepfake software to impersonate his ex-wife, significantly disrupting the court proceedings (Times of India, 2021). In another case in Nigeria, a suspect remotely impersonated a witness using a deepfake video during a corporate fraud trial, delaying the proceedings by several months. Such impersonations compromise due process, mislead judicial officials, and erode trust in virtual identities.
- **Manipulation of video evidence:** Research from the University of California, Berkeley, found that around 40% of legal professionals could not distinguish between deepfake and authentic video evidence, putting significant miscarriages of justice at risk (UC Berkeley, 2022). Furthermore, legal professionals' ethical obligation to ensure a fair trial is challenged by their limited ability to detect sophisticated synthetic content. A 2021 survey by the National Association for Court Management found that fewer than 30% of court staff had received training in digital media authentication. In 2022, a deepfake video was presented during a mock trial in the UK, where jurors accepted falsified surveillance footage as real, illustrating the deceptive potential of synthetic images.
- **Audio manipulation (voice deepfakes):** Colombian criminal groups have successfully employed voice cloning technology to impersonate law enforcement, creating fraudulent evidence or overriding authentic audio recordings, leading to misleading investigators and influencing trial outcomes (El Tiempo, 2023). Furthermore, a recent case in Canada involved a synthetic confession fabricated through cloned audio of a defendant, which was subsequently used in an extortion attempt before being discovered during forensic analysis.
- **Coercion and Blackmail:** The FBI reports an increasing use of deepfake-based blackmail targeting judicial officials and legal professionals, undermining their impartiality and integrity (FBI, 2023). In Eastern Europe, a fake deepfake video depicting a judge in a compromising situation spread on social media just days before a major corruption sentencing, sparking a public outcry and delaying the verdict.

The systemic erosion of judicial credibility, potential unjust convictions, intimidation of key stakeholders, and growing public distrust underscore the seriousness of these threats.

5. Legal and ethical implications

Deepfake technology fundamentally undermines judicial confidence in the authenticity of evidence, raising serious ethical and procedural implications. Legal systems rely on the integrity of evidence and the authenticity of witness statements, both of which are vulnerable to manipulation through synthetic media. As digital recordings become increasingly prevalent in court proceedings, the assumption that audiovisual content reflects the truth is no longer tenable.

Cases such as *People v. Smith* (California, 2022) illustrate how deepfake-related uncertainties can affect due process, particularly regarding the Sixth Amendment right to confront one's accuser. The defense successfully argued that the integrity of virtual testimony was compromised by the risk of tampering, leading to judicial review of virtual evidence policies.

The European Commission's Ethics Guidelines for Trustworthy AI (2020) emphasize the importance of human oversight, transparency, and accountability—principles that are often difficult to implement with current virtual hearing technologies. For example, synthetic media can circumvent these ethical controls when introduced through unofficial or pre-recorded testimony.

In another case, a German court temporarily suspended the use of remote testimony in a high-profile fraud case after suspicions arose that a key witness's pre-recorded video testimony had been artificially altered. This suspension highlights how the mere suspicion of deepfakes can delay justice and erode trust in the process.

In contrast, as of 2024, no international arbitration award has explicitly addressed or cited the issue of deepfakes (Just Mundi, 2025). Despite growing awareness, international arbitration has yet to encounter a case in which a tribunal has formally recognized or ruled on the authenticity concerns raised by deepfake content.

Legal experts are calling for clear evidentiary standards for synthetic media. Their suggestions include mandatory forensic validation of video and audio files, audit trails for digital presentations, and penalties for the malicious use of deepfake content. Without these reforms, courts risk systemic failures stemming from manipulated or unverifiable information.

6. Evidentiary impact

In trials, all parties involved must carefully examine the evidence. Nowadays, even evidence that appears authentic must be questioned, especially with the rise of deepfakes, due to the risk of inaccurate rulings that could lead to invalidity or even the denial of rights.[6]

This phenomenon is known as the "deepfake defense," which argues that the standard of authenticity for alleged deepfakes must be re-evaluated. John P. LaMonaga argues that

courts will have to resort to alternative methods to sufficiently determine whether photographic evidence is truly what its proponent claims.[7]

In the United States, for example, it has been pointed out that existing legal standards may be inadequate for current judicial realities. Federal Rule of Evidence 901 establishes that, to satisfy the requirement of authentication or identification of an exhibit, the proponent must present sufficient evidence to support the conclusion that the item is what the proponent claims it to be.[8]

But ultimately, this does not meet current demands, resulting in the continued use of outdated methods from years past.

7. Organized crime and judicial manipulation

Organized crime entities skillfully exploit deepfake technologies to manipulate court outcomes, evade prosecution, and disrupt legal investigations. Europol (2022) has reported cases where criminal networks used synthetic avatars and fake identities to impersonate witnesses and present false testimony during remote depositions. These deepfakes led to mistrials and the dismissal of crucial legal actions due to the inability of courts to verify the authenticity of participants remotely.

In a notable incident in Eastern Europe, a drug cartel used a deepfake version of a confidential informant to present misleading testimony via video link. The manipulation wasn't discovered until months later, after biometric inconsistencies in the video frames were compared. The mishandled testimony led to the acquittal of several defendants and the collapse of a high-profile drug prosecution.

In Brazil, investigators in Operation Car Wash reported attempts by members of criminal organizations to present falsified contracts and digitally altered audio recordings to discredit cooperating witnesses and hinder corruption investigations. Although these were ultimately detected by forensic teams, the time and judicial resources required to verify their authenticity delayed the proceedings for several months.

Cybercriminal groups have also leveraged deepfakes to impersonate law enforcement officials or lawyers. In one example from the United States, law enforcement intercepted a fake video purportedly showing a prosecutor offering leniency in exchange for bribes. Although it was fake, the video caused significant reputational damage and prompted a lengthy internal investigation.

These cases highlight the operational sophistication of modern organized crime, which is now incorporating AI capabilities into its legal interference strategies. The erosion of reliable testimony, the introduction of synthetic documentation, and the obstruction of legal accountability are profound dangers that demand an immediate technological and regulatory response.

8. Impact of AI on justice

Artificial intelligence is increasingly being integrated into judicial systems around the world, offering both transformative benefits and considerable risks.

Positive contributions of AI

AI has demonstrated enormous potential to improve the speed, accuracy, and accessibility of legal processes:

- **Improved transcription and record keeping:** Tools like JAVS (Justice AV Solutions) use AI to accurately transcribe hearings, reducing the workload on court reporters and enabling multilingual support for non-native speakers. These systems can identify speakers, tag case metadata, and create searchable files for proceedings.
- **Predictive analytics for sentencing and bail:** In jurisdictions such as the United States and the United Kingdom, AI algorithms analyze historical sentencing data to help judges make more consistent, data-driven decisions. These tools, when properly calibrated, reduce disparities and promote equitable justice.
- **Translation and Accessibility:** AI-based translation software and real-time language processing improve accessibility for people with limited language proficiency, increasing their ability to meaningfully participate in court proceedings. This is exemplified by platforms like Google Meet, which offer instant speech translation to meet this need.
- **Detecting Judicial Bias and Corruption:** In Brazil, Operation Lava Jato used AI systems to uncover patterns of corruption by cross-referencing large volumes of contracts, legal documents, and communications, leading to numerous convictions and institutional reforms.

Risks and negative implications

- **Sole reliance on video conferencing for court records:** Courts that use video conferencing platforms as their sole source for capturing and preserving court records risk introducing a single point of failure. Without additional information security safeguards, such as biometric verification, blockchain integrity controls, or record-keeping redundancy, manipulated video feeds—whether through deepfake visual impersonation or altered audio—could be recorded as authentic proceedings. For example, in 2021, a virtual small claims hearing in California was invalidated when it was discovered that the recording platform had been falsified to replace one party's feed with a pre-recorded deepfake statement. The official transcript relied on this manipulated recording, raising concerns about digital chain of custody and authentication protocols.
- **Deepfake Threat Rising Due to Improper AI Implementation:** One of the most insidious risks arises when courts adopt AI systems without rigorous scrutiny, inadvertently introducing flawed or insecure technologies that can be exploited by malicious actors. For example, AI-based identity verification tools with poor

security may fail to detect sophisticated deepfakes, allowing imposters to provide false testimony or present falsified evidence. In one such case, a criminal group exploited a procurement flaw in a low-level identity verification API to access court records and impersonate attorneys in filing fraudulent legal motions. By using deepfake content in combination with poorly implemented AI, these actors effectively created an alternative judicial narrative, misleading the court and delaying legal resolution for months.

Despite these advantages, AI also presents serious dangers to justice if misused or deployed without oversight:

- **Algorithmic bias and discrimination:** AI models trained on biased historical data can replicate and even amplify existing inequalities. For example, some pretrial risk assessment tools have shown racial and socioeconomic biases, particularly in predicting recidivism.
- **Overreliance on automation:** Judges and lawyers may overrely on algorithmic recommendations without fully understanding the model's limitations, which can lead to unfair outcomes. Decision-making risk is particularly concerning when AI systems lack transparency and accountability.
- **Evidence tampering and falsification:** Criminals can use generative AI to produce synthetic evidence, such as fake audio or altered documents. Without robust validation protocols, courts risk accepting fraudulent material as legitimate.
- **Data privacy and security:** The increasing digitization and centralization of sensitive judicial data make judicial systems prime targets for cyberattacks. AI systems must be reinforced with robust data governance policies to prevent breaches and misuse.
- **Loss of human judgment:** Justice involves empathy, discretion, and moral reasoning—qualities that AI cannot replicate. There is a danger that overreliance on AI tools will erode these essential human aspects of judicial decision-making.
- **Open Access:** Deepfake technology is becoming increasingly accessible, with many applications available for free and even operable on mobile devices. The widespread availability of open-source machine learning algorithms and the ability to run them on everyday hardware make these tools easy to obtain and use. However, there is little to no oversight over who accesses or uses these systems, raising serious concerns, as it allows malicious actors to exploit this technology with minimal barriers or accountability.[9]

In short, while AI offers extraordinary tools to improve the administration of justice, its implementation must be carefully regulated in both the public and private sectors, especially when sensitive information is involved. Systems must be transparent, explainable, ethically designed, and subject to rigorous oversight to avoid miscarriages of justice.

9. Solutions and safeguards

To address the growing threat of deepfakes in legal proceedings, a multi-pronged approach combining technological, procedural, and legal strategies is essential:

- **Deepfake Detection with BioID:** BioID offers robust facial recognition systems integrated with deepfake detection algorithms capable of identifying anomalies in synthetic media. These tools use AI models trained to detect inconsistencies in facial muscle movement, eye reflection, and frame continuity, providing an essential layer of authentication during remote hearings.
- **Liveness Detection:** A key feature of BioID technology is liveness detection, which ensures that the person presenting the credentials is physically present and not a digital forgery. This real-time verification can prevent imposters from submitting testimonials using manipulated video or audio.
- **Facial recognition and biometric verification:** Facial recognition platforms can compare participants with stored biometric data, ensuring that witnesses, defendants, and court personnel are who they say they are. Courts can integrate this with multi-factor authentication systems to more effectively protect remote identities.
- **JAVS High-Fidelity Recording and Real-Time Analysis:** JAVS courtroom systems offer high-resolution audio and video recording capabilities with advanced speaker diarization technology. This feature accurately attributes speech to a participant, mitigating misidentification or tampering. Additionally, JAVS supports real-time transcription and translation, enabling multilingual access and improving the transparency of proceedings for non-native speakers.
- **Redek ODR Platform :** ODR (Online Dispute Resolution) is the use of digital platforms to resolve disputes between parties through virtual hearings or other online proceedings, without the need for in-person appearances. It offers a secure, efficient, and accessible way to manage disputes. Redek's AI-powered ODR platform provides a reliable and structured environment for managing virtual court proceedings with integrity and transparency. Unlike basic video conferencing tools, Redek functions as a complete digital repository that records, timestamps, and audits all procedural actions, whether document submission, editing, or system interaction. This architecture ensures full traceability and prevents unauthorized modifications, as any attempt to insert synthetic or manipulated content would lack the corresponding metadata and disrupt the document lifecycle.

The system is specifically designed to detect inconsistencies and preserve the authenticity of digital files. For example, a false or falsified submission would be immediately detected due to the lack of a traceable record within the platform. This improves procedural security and mitigates the risk of evidence tampering.

- **Blockchain-based evidence management:** By recording audio, video, and text evidence on an immutable blockchain ledger, courts can maintain a verifiable chain of custody and protect against post-submission tampering.
- **Training and capacity building:** Legal professionals must receive specialized training to recognize and respond to deepfakes. This includes knowledge of forensic analysis tools, protocols for authenticating evidence, and awareness of legal standards relating to AI-generated content.
- **Legislative and regulatory frameworks:** Governments should update evidentiary standards to define admissibility criteria for digital submissions and establish penalties for the creation or use of deepfakes to obstruct justice. Regulatory bodies should also certify AI tools used in courts to ensure compliance with ethical and procedural standards.

10. Physical presence

While technological advances have considerably improved judicial systems, physical presence in courtrooms remains essential to minimize vulnerabilities to digital manipulation. Furthermore, in-person hearings have been shown to help judges better understand the nuances of a case. Direct, in-person verification methods ensure the authenticity of testimony and identity. The landmark ruling in *Crawford v. Washington*, 541 U.S. 36 (2004), reinforces the need for face-to-face confrontation to maintain judicial integrity and procedural fairness. The Supreme Court emphasized that the Sixth Amendment's Confrontation Clause guarantees the defendant's right to confront witnesses, a right that is significantly undermined in remote or virtual formats.

International legal frameworks also uphold the right to physical presence. Article 14 of the International Covenant on Civil and Political Rights (ICCPR) affirms the right to a fair trial, including the accused's ability to be present and examine witnesses. The European Court of Human Rights has similarly ruled that remote participation should not impair the accused's ability to effectively participate in the proceedings (e.g., *Marcello Viola v. Italy*, 2019).

Numerous courts have reaffirmed the need for in-person proceedings after incidents in which remote formats failed to prevent tampering. In a 2021 case in Ontario, Canada, a judge overturned a conviction after finding that the defendant had been digitally replaced during part of the virtual trial. The court held that virtual hearings should not compromise fundamental legal rights and ordered stricter in-person safeguards for future cases.

In-person trials also offer subtle but crucial nonverbal cues, such as demeanor, body language, and courtroom conduct, that influence judicial perception and decision-making. These cues are difficult, if not impossible, to accurately assess in digital formats, especially when deepfake technology can simulate such expressions.

Therefore, while virtual technologies can complement judicial proceedings in emergencies or logistical situations, the fundamental principles of justice and due

process still depend on the irreplaceable protections afforded by physical presence in the courtroom.

11. Challenges and dangers for judicial systems

The integration of AI and deepfake technologies into judicial systems, while potentially beneficial, also presents a number of pressing challenges and dangers that require urgent and comprehensive intervention.

- **Authentication Failures:** Deepfake videos and audio can circumvent current identity verification mechanisms, especially in remote hearings. In 2022, a Florida district court mistakenly accepted a falsified videotaped deposition due to inadequate biometric security measures, resulting in a mistrial after the deception was discovered during appellate review (Florida Judicial Review Board, 2023).
- **Admissibility Ambiguities:** Legal frameworks across jurisdictions lack clear rules for determining the admissibility of AI-generated or potentially synthetic evidence. In the UK, a 2021 Crown Court case involving altered video evidence led the investigating judge, Mr. J, to expand pre-trial arguments on admissibility and the burden of proof in relation to the detection of deepfakes, significantly delaying the court's ruling.
- **Erosion of public trust:** A 2022 survey by the International Bar Association found that 62% of respondents expressed less confidence in virtual court proceedings due to perceived vulnerability to evidence tampering. The mere possibility of deepfakes can lead to public skepticism about the legitimacy of verdicts and the fairness of trials.
- **Resource constraints:** Deepfake detection tools and AI-based authentication technologies require significant investment in infrastructure and staff training. Developing countries or under-resourced judicial districts may struggle to acquire or maintain the necessary technical capabilities, exacerbating inequalities in access to justice.
- **Weaponization by malicious actors:** Cybercriminals and politically motivated groups can use deepfakes to attack judicial officials, fabricate public scandals, or distort testimony. A high-profile case in South Africa involved a fake audio recording of a judge allegedly accepting bribes, which circulated widely before being debunked by audio forensics (Johannesburg Legal Monitor, 2022).
- **Platform fragmentation and lack of standardization :** Another pressing concern is the multiplicity of platforms used to conduct hearings and manage court files. Judicial institutions often rely on fragmented systems—email chains, insecure cloud storage, siloed case management tools—that undermine the traceability, consistency, and interoperability of information. This lack of standardization complicates oversight, creates data silos, and increases the risk of deepfakes.

To effectively respond to these threats, judicial systems must adopt a multifaceted defense strategy, combining AI tools with human oversight, real-time biometric verification, interagency cooperation, and updated legal protocols. **Online dispute**

resolution (ODR) platforms offer a centralized and standardized digital environment that improves data integrity, procedural transparency, and information security.[10]

Conclusion

Judicial integrity in the face of rapidly evolving digital threats depends on proactive technological integration, strict legal frameworks, enhanced verification mechanisms, and the maintenance of essential court traditions. Addressing these challenges through a balanced approach ensures judicial fairness, credibility, and public trust.

However, the selection of technological tools must be approached with extreme caution. The deployment of poorly tested or poorly vetted AI systems, particularly those responsible for identity verification or evidence analysis, can open the door to the exploitation of deepfakes, allowing malicious actors to easily manipulate proceedings. In these cases, courts risk becoming vehicles for false narratives constructed from AI-generated content, from manipulated testimonies to falsified audiovisual evidence.

These vulnerabilities not only threaten the integrity of individual trials but also undermine the fundamental principles of due process, fairness, and transparency. The long-term consequence is a weakened judiciary susceptible to external manipulation, which in turn poses a profound threat to the rule of law and democratic governance itself. If justice can be synthetically altered, public confidence in the legitimacy of democratic institutions is eroded, fostering cynicism, disengagement, and social instability.

Therefore, protecting the judicial system from threats related to deepfakes and AI is not just a technical or legal necessity, but a democratic imperative.

---

[1]Nicolás Ernesto Lozada Pimiento is the CEO of REDEK, the first LegalTech solutions platform in Latin America, recognized for its innovation in digital justice and online dispute resolution. He also works as a professor at the Externado University of Colombia, where he teaches business law, with an emphasis on arbitration and legal technology.

[2]Ed Segura is a senior executive specializing in the digital transformation of judicial systems. He leads innovation in legaltech through the use of AI, biometrics, and data analytics. With deep expertise in emerging technologies, he drives efficiency, transparency, and reform in judicial ecosystems.

[3]Kaur, A., Noori Hoshyar, A., Saikrishna, V., Firmin, S., & Xia, F. (2024). Deepfake video detection: challenges and opportunities. Artificial Intelligence Review, 57(6), 159

[4]Mashable. (2025). Google Veo 3 AI video tool floods the internet with real-looking clips. Mashable.https://mashable.com/article/google-veo-3-ai-video

[5]Laux, J., & Kröger, M. (2022). Video conferencing in proceedings. In H. Ruiz Fabri (Ed.), Max Planck Encyclopedia of International Procedural Law. Oxford University Press.

[6]The virtual courtroom at risk: Deepfake technology and the undermining of judicial integrity (2024).

[7]Cornell Law School. (n.d.). Rule 901. Authenticating or identifying evidence. Legal Information Institute.https://www.law.cornell.edu/rules/fre/rule_901

[8]Dixon Jr., H. B. (2024). The “Deepfake Defense”: An evidentiary conundrum. The Judges' Journal, 63(2), 38–40. American Bar Association.

[9]Vaswani, A., & Shazeer, N. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.https://arxiv.org/abs/1802.07228v2

[10]Redek's platform consolidates case management, biometric verification, algorithmic resolution pathways, and real-time analytics into a unified, cloud-based system that is ISO-certified and API-compatible. It supports scalable automation and can be configured for a variety of dispute types, from small claims to family law and consumer arbitration.