Deepfake Evidence in Arbitration the Upcoming Challenge for Digital Justice

Mr. Rishabh Mishra
3rd Year B.A LL. B (Hons.) Student At Indore Institute of Law, Indore

***Abstract*—The rapid growth of Artificial Intelligence (AI) has changed the generation, preservation, and sharing of digital data, presenting both new prospects and legal challenges. One of the most notable innovations in this regard is the advent of deepfake technologies which involve synthetic audio, video, and images created by means of AI that can accurately mimic real people and situations. However, apart from having many beneficial applications in such spheres as education, entertainment, and innovation, deepfakes represent a significant threat to the reliability of digital evidence which is vital to modern dispute resolution processes.**

Deepfake technologies have become increasingly worrying to many. For example, the Global Risks Report 2025 published by the World Economic Forum lists AI-generated misinformation and disinformation among the top global risks of a shorter term while UNESCO's Guidelines for the Use of AI Systems in Courts and Tribunals (2024) stress the importance of transparency, human oversight, and necessary precautions to maintain the credibility of judicial and quasi-judicial proceedings. Moreover, in India, the Ministry of Electronics and Information Technology (MeitY) has provided relevant guidelines concerning the role of digital intermediaries in preventing the spread of deceptive AI-generated content.

The aforementioned issues acquire importance especially in cases of arbitration, as the process more and more depends upon electronic agreements, e-mails, digital communications, and virtual meetings. While the adaptability of the arbitral procedure remains its most valuable asset, it also subjects arbitrators to the threat of advanced AI-generated evidence which might be hard to spot by traditional means. With technology developing much faster than law, the question facing arbitration is whether it will be possible for the current standard of proof to reliably separate real digital evidence from AI-based deception.

***Index Terms*—Deepfake Evidence; Arbitration; Digital Justice**

I. THE TECHNOLOGY OF DEEPFAKE AND ITS LEGAL IMPLICATION

Deepfake technology is a term that is derived from the fusion of "deep learning" and "fake" and describes media that is artificially produced by the use of highly developed artificial intelligence technology like deep neural networks and Generative Adversarial Networks (GANs). Such technology works through analyzing huge amounts of data sets of images, videos, and recordings of voices in order to produce highly realistic material that can replicate someone's expressions, way of speaking, and other manners. From being used only in research laboratories, this technology is now easily accessible through commercial products as well as free of charge AI tools.

Legal issues connected with this technology do not only include the issues of invasion of privacy and spreading disinformation in the online environment. Today, in disputes, digital materials such as emails, video materials, recordings from surveillance cameras, voice messages, and recordings from virtual meetings can often serve as a basis for claims about some facts. The introduction of AI technology that produces media fundamentally transforms the established presumption that visual or audio records are always authentic.

Given the above dangers, organizations like UNESCO and World Economic Forum have cautioned that there is a danger that the use of generative AI could lead to the erosion of trust in the information environment as well as the legal processes involved. In arbitration, which involves procedural flexibility as well as heavy use of electronic evidence, the issue of deepfake technology poses a great threat to the integrity of the process.

II. EVIDENCE FRAMEWORK: CAN ARBITRATION CATCH UP WITH AI-BASED EVIDENCE?

The increasing digitization of commercial dealings has completely changed the kind of evidence that is produced before arbitration tribunals. E-mails, e-contracts, WhatsApp messages, cloud data, online invoices, closed-circuit televisions, and audio-visual recordings from video conferencing are essential tools for proving the existence of the contract and commercial disputes. In order to cater to the changing times, the Arbitration and Conciliation Act, 1996 has provided a flexible attitude towards the rules of evidence. Under Section 19 of the said act, it is clearly mentioned that the arbitral tribunals are not governed by the Code of Civil Procedure, 1908 or the Indian Evidence Act, 1872.

Even though the evidentiary process is not guided by specific procedural requirements in arbitration, the genuineness and veracity of the electronic record are vital issues. According to the Information Technology Act, 2000, electronic records and electronic signatures were made legal, and the Supreme Court of India in cases Anvar P.V. v. P.K. Basheer (2014) and Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal (2020) highlighted the need for proving the genuineness of the electronic evidence through legislation. In international practice, the IBA Rules on the Taking of Evidence in International Arbitration (2020) and the UNCITRAL Model Law on

International Commercial Arbitration support the flexibility of the process in taking evidence but do not address the issue of difficulties associated with AI-generated synthetic media. The development of deepfake technology demonstrates the deficiency in regulation.

III. CHALLENGES TO DIGITAL TRUST: DEEPFAKE EVIDENCE AND THE FUTURE OF ARBITRATION

The ever-changing nature of the technology of deepfaking presents a multi-faceted threat to arbitration, going well beyond technical issues to include the very basis of evidentiary authenticity and procedural justice. The modern-day arbitration process makes use of electronic evidence, such as emails, instant messages, electronic contracts, security camera footage, and audio-video evidence, along with the transcripts of the hearings held online, and all this is vulnerable to being tampered with using AI. Whereas usual digital manipulation is easily detected, deepfakes are capable of imitating someone's voice, facial expressions, and mannerisms, making any doctored content hard to spot without proper forensics. This brings into question the concepts of authenticity and reliability of electronic evidence, upheld by the Supreme Court of India in the cases of Anvar P.V. v. P.K. Basheer (2014) and Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal (2020). While, on the one hand, Section 19 of the Arbitration and Conciliation Act, 1996 allows for some flexibility on the part of the arbitral tribunal, on the other hand, it increases their burden of proof as far as disputed digital evidence goes. In the international arena, both the IBA Rules on the Taking of Evidence in International Arbitration (2020) and the UNCITRAL Model Law on International Commercial Arbitration have not addressed the use of AI to create synthetic media, which is a notable gap in regulation. In addition, the involvement of digital forensics experts and the process of authenticating such evidence may add up to costs and length of the process of arbitration, thus going against its fundamental principles of efficiency and economy. Taking into account the above problems, the Guidelines for the Use of AI Systems in Courts and Tribunals issued by UNESCO (2024) require transparency and human oversight in decision-making, while the World Economic Forum highlights AI-generated misinformation among the greatest short-term global risks (Global Risks Report 2025). The above illustrates that the issue related to deepfakes cannot be viewed only through technological lenses – it undermines the very trust that the arbitration process and digital evidence are based on.

IV. COMPARISON OF INTERNATIONAL DEVELOPMENTS: LESSONS FOR THE FUTURE OF ARBITRATION

The accelerated development of artificial intelligence technology has led to the introduction of regulatory measures designed to protect the integrity of legal proceedings. The Artificial Intelligence Act of the European Union (Regulation (EU) 2024/1689) is an example of a comprehensive legislative framework in terms of the protection of judicial proceedings, as it defines AI systems in use for legal procedures and alternative dispute resolution as high-risk

systems, thus requiring compliance with the stringent standards concerning transparency, human supervision, accuracy, and accountability.

In the international arena, the United Nations Educational, Scientific and Cultural Organization has issued the first global ethics guidelines for AI systems in courts and tribunals (2024). Drafted after consultations of over 160 experts from more than 160 countries and over 36,000 judicial stakeholders, the guidelines focus on the principles of transparency, accountability, and human oversight. In a similar fashion, the European Commission for the Efficiency of Justice (CEPEJ), which is a statutory body of the Council of Europe, calls for the utilization of AI in judicial reasoning while retaining judicial discretion, and highlights the importance of data quality, cybersecurity, and non-discrimination.

International arbitration, however, is yet to have an explicit legal framework on the use of evidence created by AI or deepfakes. Even though the UNCITRAL Model Law and the IBA Rules on the Taking of Evidence in International Arbitration 2020 give arbitral tribunals wide powers in the assessment of evidence, there is still no provision concerning the authenticity of artificial media. The foregoing deficiency proves that whereas legal systems all over the world are gradually adjusting to artificial intelligence, international arbitration requires further development in this field too.

V. REINFORCING ARBITRATION IN RESPONSE TO DEEPFAKE EVIDENCE: THE PATH AHEAD

In light of the evolving nature of deepfake technology, the mechanism of arbitration should take the lead in devising necessary safeguards before disputes over digital evidence can arise. Even though Section 19 of the Arbitration and Conciliation Act, 1996 gives wide powers to arbitral tribunals regarding their procedures, such discretion should be used in formulating necessary steps to ensure that the authenticity of such digital evidence is confirmed through proper verification. Such verification process will include making it mandatory for disputing parties to provide information regarding the origin, format, and chain of custody of such digital records as well as other means such as metadata, hash values, among others, that may help in establishing the integrity of such electronic evidence.

Institutional reform is just as important. It should be explored whether the leading arbitral institutions, such as ICC, SIAC, LCIA, and MCIA, would want to implement AI-specific procedures for the disclosure, authentication and forensics of the AI-generated evidence. The existing institutional practices are flexible enough as they incorporate the rules of the IBA Rules on the Taking of Evidence in International Arbitration (2020), whereas the UNCITRAL Model Law could serve as the base for creating internationally harmonized standards on synthetic digital evidence. Such institutional reforms correspond to the UNESCO Guidelines for the Use of AI Systems in Courts and Tribunals (2024) and their emphasis on human supervision, transparency and accountability of AI-assisted adjudication processes. Finally, confidence in arbitration will not

only require embracing of technological advancements, but creation of the evidentiary procedure that will guarantee that digital efficiency will never come at the cost of truth and justice.

VI. CONCLUSION

Institutional reform is similarly important. Institutions like ICC, SIAC, LCIA, and MCIA ought to contemplate incorporating specific protocols for the use of AI-based evidence in respect of its identification, validation, and examination. This will go hand in hand with the IBA Rules on the Taking of Evidence in International Arbitration (2020), while the Model Law can be the base of international standards. As per the Guidelines for the Use of AI Systems in Courts and Tribunals by UNESCO (2024), these reforms will improve transparency, human intervention, and accountability, which are necessary to ensure that technology improves arbitral procedures.