Journal of Intellectual Property and Information Technology Law (JIPIT)
ISSN: 2788-6727 – Volume 5, Issue 1 (2025) p.105-133
DOI: https:doi.org/10.52907/jipit.v5i1.624

Strengthening the Kenyan Court-Annexed Mediation Against the Threat of Deepfakes and Digital Deception

Raphael M. Okochil[*]

ABSTRACT

This paper explores how Kenya's mediation framework, particularly the Court-Annexed Mediation (CAM) system, can adapt to an era where fabricated evidence (such as Artificial Intelligence (AI) generated deepfakes and distorted digital documents) threatens the principles of truth, trust, and fairness, which are essential elements in Alternative Dispute Resolution (ADR). This analysis focuses on mediation due to its distinctive vulnerabilities: its inherent informality, the absence of formal evidentiary rules, and the mediator's non-adjudicative role, all of which make it exceptionally susceptible to manipulation by synthetic media. Relying on Kenya's Civil Procedure (Court-Annexed Mediation) Rules, 2022, the Evidence Act (Cap. 80), and emerging local jurisprudence, this paper examines how the law treats digital and synthetic evidence in ADR. Further, this paper relies on interdisciplinary literature from cognitive psychology and digital ethics to frame the challenge. Thus, the author addresses several sub-themes, such as the 'illusion of truth', the cognitive tendency to believe convincing fakes, and the ethical and procedural challenges that arise when fabricated information enters the dispute resolution process. Additionally, the author explores the promises and shortcomings of technologies such as AI-assisted tools for detecting such forgeries, considering their practical (in)accessibility to Kenyan mediators. To this end, this paper argues that the future of mediation in Kenya requires mediators to become more vigilant facilitators, armed with digital awareness and supported by frameworks that enable them to broker peace in an era where appearances no longer guarantee reality. This paper proposes a set of integrated reforms by examining the evolving nature of evidence and human judgment within the facilitative process. It also proposes institutional safeguards to be deployed when handling digital materials.

* Raphael Okochil is the founder of the ADR & the Law Journal. He enjoys legal research and writing and has published articles in reputable journals. Email: raphaelokochil@gmail.com

Keywords: Artificial Intelligence, Deepfakes, Digital Ethics, Mediation, Synthetic Evidence

TABLE OF CONTENTS

I. INTRODUCTION

Kenya's Judiciary has progressively institutionalized Alternative Dispute Resolution (ADR) through its Court-Annexed Mediation (CAM) program, which concluded four thousand four hundred and fifty-one cases in the 2022/2023 financial year alone, marking a ninety-five percent conclusion rate (Judiciary of Kenya, 2023, p. 138). This success depends upon a trust-based, informal process designed to encourage open dialogue and party autonomy. However, this very informality is now threatened by the proliferation of synthetic media, including AI-generated videos and audio, which can convincingly forge evidence and manipulate the truth, thereby challenging the foundational pillars of fairness in ADR.

To situate this analysis, it is essential first to define mediation. The Civil Procedure (Court-Annexed Mediation) Rules of Kenya, 2022 (hereinafter the CAM Rules) define mediation as an 'informal and non-adversarial process conducted physically or virtually where a mediator encourages and facilitates the resolution of a dispute between two or more parties' (Civil Procedure (Court-Annexed Mediation) Rules, 2022, r. 2). This process typically begins when a court, either on its own motion or at the request of the parties, refers a suitable case for mediation, often after a screening process (Civil Procedure (Court-Annexed Mediation) Rules, 2022, rr. 5, 6). The parties then, with the assistance of a court-accredited mediator, engage in discussions through joint sessions (plenary) and private meetings (caucuses) (Nyaga, Ondego & Joel, 2023, p. 3). The mediator's role is not to impose a decision but to guide communication, identify underlying interests, and assist the parties in crafting their own mutually acceptable settlement agreement (Sessional Paper No. 4 of 2024 on the National ADR Policy, p. 34). This facilitative, party-driven, and confidential nature is mediation's greatest strength, but it is also its greatest vulnerability in the digital age. This paper argues that without anticipatory legal and professional reforms, Kenya's CAM system is procedurally unequipped to handle dis-

putes tainted by synthetic evidence, risking the erosion of party autonomy and public trust in ADR.

This crisis is not so much technological as human. The challenge posed by synthetic media is fundamentally human because it exploits the innate cognitive biases and the trust-based architecture of human social interactions (Kahneman, 2011, p. 4). Technology is merely the tool; the true vulnerability lies in human psychological wiring, which makes humans susceptible to well-crafted deception (Nyaga, Ondego & Joel, 2023, p. 7). The crisis, therefore, demands a response that is not just technologically sophisticated but also deeply attuned to the human element of dispute resolution.

While all forms of dispute resolution are affected, mediation relies on an assumption that people come to the table with information that can be trusted (Bush & Folger, 2005). The process is prized for its key attributes, including party autonomy, confidentiality, flexibility, and its ability to preserve relationships, making it a cornerstone of the ADR mechanisms promoted under Article 159(2)(c) of the Constitution of Kenya, 2010 (Barasa, 2025, p. 2). Mediation assumes, even in its informality, a shared commitment to fairness (Maute, 1990, p. 363).

This paper argues that mediation is vulnerable to the threat of deepfakes for several reasons that distinguish it from other ADR mechanisms, such as arbitration. First, mediation operates in a space where it intentionally lacks the formal rules of evidence that govern litigation and, to a lesser extent, arbitration (Stempel 1997, p. 961). In Kenya, for instance, the Evidence Act (Cap. 80) is explicitly disapplied from adjudication proceedings under the proposed Construction Adjudication Bill, and a similar informality characterizes mediation practice, creating a procedural vacuum where fabricated evidence can thrive (Sessional Paper No. 4 of 2024 on the National ADR Policy, p. 117). This informality, designed to foster open dialogue, becomes a critical flaw when confronted with sophisticated forgeries, as there is no established mechanism for admissibility challenges or forensic scrutiny (Prigoff, 1988, p. 2).

Second, the mediator's role is purely facilitative, not adjudicative; they are not empowered to act as a fact-finder, weigh evidence, or determine truth (Stempel 1997, p. 950). The Kenyan CAM Rules reinforce this by defining the mediator's role as one who 'encourages and facilitates' rather than decides (Civil Procedure (Court-Annexed Mediation) Rules, 2022, r. 2). Their primary function is to manage the conversation (Stempel, 1997, p. 961), which can be dangerously skewed by compelling but fabricated material. Finally, mediation often centers on emotional narratives and interpersonal dynamics (Williams & Hinshaw, 2018, pp. 165, 167), the very domains that hyper-realistic deepfakes are designed to exploit (Qureshi & Khan, 2024, pp. 102-103), making both parties and mediators susceptible to manipulation that a more formal, document-focused process might mitigate.

At the same time, the legal environment had not prepared for this storm. While recent studies highlight the significant threat deepfakes pose to the formal criminal justice system by compromising evidence and eroding trust (Sandoval *et al*., 2024, p. 1), it is logical to infer that most ADR frameworks, which often operate with even less stringent evidentiary rules, have yet to appreciate the magnitude of these challenges fully. Kenya's legal system, for example, has developed a rigorous framework for authenticating digital evidence in court, requiring forensic soundness, expert testimony, and adherence to Sections 106A and 106B of the Evidence Act (Cap. 80) to ensure reliability (Rutenberg, Kiptinness & Sugow, 2021, p. 36).

In the event that parties tender AI-fabricated evidence, the mediation process itself is cast into dangerous waters. The mediator, who is expected to facilitate the process rather than render a determination (Stempel, 1997, p. 950), will face the challenge of guiding the dialogue when the very sensory inputs are unreliable (Citron & Chesney, 2019, p. 1785). The digital age has ushered in what some have called the 'post-truth' era (Keyes, 2004). As philosopher Arendt (1973) warned, 'the ideal subject of totalitarian rule is not the convinced Nazi or the convinced

Communist, but people for whom the distinction between fact and fiction… no longer exists' (p. 474).

Additionally, if a settlement is reached based on fabricated evidence, what recourse exists? This question is less about holding the mediator liable (since they are rightly shielded from liability that is devoid of bad faith) (Schulz, 2023) and more about the integrity of the outcome. It is about the procedural care standard, if any, that applies to the process in such instances. The law remains vague, but the consequences are not. A settlement born of deception ruins lives, damages trust in the system, and renders the promise of justice useless (Kakulungulu 2025). 'Law without justice is a wound without cure', said Scott Downey (JusCorpus, 2023). In this digital age, that wound is deeper than ever and challenging to see.

Technology, of course, promises to offer solutions. AI can be used to verify the authenticity of an audio or video file (Verdoliva, 2020, p. 8). Additionally, technology is evolving to flag inconsistencies in speech patterns and detect the subtle glitches in videos that often accompany deepfakes (Verdoliva, 2020, p. 10). However, these tools are not a panacea and can create an over-reliance on technology in a process that was designed to be human-centered (Gillett, 2025). Furthermore, in Kenya, the availability, cost, and technical expertise required to use such tools are significant barriers. Most mediators accredited by the Judiciary's Mediation Accreditation Committee (MAC) are legal or social science professionals, not digital forensic experts, and their standard training does not cover the use of AI-powered detection tools (Barasa, 2025, p. 6).

Therefore, this paper returns to the mediator: not as a passive facilitator, but as an active guide in a new environment. This does not mean that they must embody the roles of detective or judge, but that they must be more vigilant managers of the process. The mediator of the future will have to be a listener and a sceptic, a healer and a guide. Neutrality, which was once understood simply as distance from the parties, must now also en-

compass the capacity to maintain a fair and balanced process in the face of potential digital manipulation (Gaffney, 2022, p. 63). The neutral cannot afford to be naïve. They must understand the limits of their own perception, the dangers of manipulation, and the ethical dilemmas they must help the parties navigate in each mediation session.

To logically structure this inquiry, the paper is organized into six parts. Part I is the introduction. Part II delves into the cognitive science behind the 'illusion of truth', establishing the socio-legal framework for why humans are susceptible to deception by synthetic media. Part III navigates the procedural and ethical minefields that deepfakes create within the existing Kenyan mediation framework, highlighting the gap between legal principles and technological reality. Part IV examines the potential and pitfalls of using AI-powered tools to detect deepfakes, assessing whether technology can solve the problem it created. Part V redefines the concept of mediator neutrality for the digital age, arguing for a shift from passive impartiality to active, critical vigilance. It also proposes a concrete way forward, outlining specific legislative and professional reforms needed to codify digital vigilance. Finally, Part VI concludes by reflecting on the evolving nature of trust in a synthetic world.

II. A SOCIO-LEGAL FRAMEWORK FOR UNDERSTANDING DECEPTION

This paper begins by investigating why humans are susceptible to falsehood, proposing that false memories are an 'inevitable by-product of otherwise adaptive cognitive processes' like categorization (Hunt & Chittka, 2014, p. 2). This leaves individuals at the mercy of deception, especially when it is packaged in the familiar cadence of speech or the grain of a video (Graber-Mitchell, 2021, pp. 5-6). Even though professionally accredited mediators are often trained to be neutral and observant (Cloke, 2017, pp. 84, 91), they, like the parties themselves, are not immune to

deception (Cloke, 2017, p. 227). In fact, the very skills that make someone a good mediator (empathy, attentiveness, and human connection) may also make them vulnerable to deepfake manipulation (Cloke, 2017, pp. 9, 141, 146). As media scholar Boyd (2017) notes, 'the more convincing the lie, the more truth feels inadequate.' This makes it difficult for one to facilitate resolution when the emotional weight of a lie presented by one party feels stronger to the other than the fragile light of their own truth.

Therefore, it is important to establish a foundational understanding of the cognitive vulnerabilities that synthetic media exploits before delving into the specific legal and procedural problems facing mediation in Kenya. This is more than a psychological detour, since it serves as the basis for a socio-legal analysis. Cognitive psychology, which studies fundamental mental processes such as perception, memory, and judgement, is essential for understanding why individuals, including trained professionals like mediators, are susceptible to believing falsehoods, especially when presented in a compelling format (Gilbert, Malone & Krull, 1990). Understanding these inherent issues is essential for designing adequate legal and procedural safeguards.

To understand these vulnerabilities, it is essential first to define the term 'cognitive biases'. As Daniel Kahneman explains in his seminal work, *Thinking, Fast and Slow*, the human mind operates through two modes of thought: 'System 1', which is fast, intuitive, and automatic, and 'System 2', which is slower, more deliberate, and analytical (Kahneman, 2011, p. 21). Cognitive biases are systematic patterns of deviation from norms or rationality in judgment that occur because of human intuition. System 1 often relies on mental shortcuts, or 'heuristics', to make quick assessments (Kahneman, 2011, p. 12). These shortcuts are efficient but can lead to predictable errors or biases, especially when System 2, the human lazy controller, fails to engage and correct them (Kahneman, 2011, p. 49). It is these built-in mental errors that synthetic media are designed to exploit.

As Kahneman (2011) explains, most of one's daily decisions (such as decisions about the authenticity of specific evidence) are made using System 1 (Kahneman, 2011, p. 21). This is what deepfakes, synthetic tools, and manipulated documents all aim to achieve. They are very smooth, familiar, and filled with emotions (Kahneman, 2011, p. 60). They just feel right. They demand little to no scrutiny. Once System 1 has made a decision, System 2 rarely intervenes unless it is prompted by something that is evidently off (Kahneman, 2011, p. 24). In mediation, where time is limited and trust is paramount, parties and even the mediator are unlikely to engage System 2 rigorously. For a mediator in a busy Nairobi CAM session, operating under the 60-day deadline (Civil Procedure (Court-Annexed Mediation) Rules, 2022, r. 37) and handling multiple cases a day, the temptation to rely on the rapid, intuitive judgments of System 1 is immense. The cognitive load makes the deliberate, analytical scrutiny of System 2 a scarce resource.

Moreover, social psychology reveals a phenomenon called fluency heuristic: people are more likely to believe information that is easy to process (Hertwig, Herzog & Schooler, 2008, p. 1191). The 'seeing is believing' heuristic is etched into the human psyche. Studies such as 'The Camera Never Lies' by Najdowski and Stevenson (2018) show that video evidence often sways jurors in court, even when it is irrelevant to the case (p. 4). This is dangerous in a world where fake content is even more 'fluent' than real content (Citron & Chesney, 2019, p. 1767). A forged document will not contain the typos and idiosyncrasies of a tired employee. Ironically, this close-to-perfect content is more believable than the messiness of human-generated content (Laurier *et al*., 2024; Hao, 2024; Chowdhury & Lubna, 2020). Mediation, which relies primarily on personal accounts, is more likely to fall victim to this risk (Coben, Fox & Love, 2018, p. 15). If a party speaks with less fluency (perhaps due to trauma, language barriers, or nervousness), it may appear less credible than a fake video or voice note that glides smoothly through the airwaves (Citron & Chesney, 2019, pp. 1759-1760).

To err is human, but to be fooled by a deepfake is modern. In a time when pixels deceive more convincingly than human beings, the ordinary citizen (and indeed the average mediator) has little defense against a well-staged deception (Westerlund, 2019, pp. 1-4). Cognitive psychology has long shown that human beings are wired to prefer what they can easily perceive. This phenomenon, known as 'perceptual fluency', demonstrates that information that feels 'smooth' or 'natural' to process is judged more positively and becomes more emotionally appealing (Reber *et al*., 1999, p. 45). This innate preference for fluent information creates a vulnerability, as the positive feeling it generates can be easily mistaken for a sign of truth or reliability. It is not an error, per se; it is an evolutionary design survival mechanism. But that cognitive environment, shaped for the needs of hunter-gatherer groups, now operates in a digital world where algorithms and auditory experience can generate visual information that can be artificially recreated (Graber-Mitchell, 2021, pp. 2-3). Unlike face-to-face lying, which is mental and can be betrayed by non-verbal cues associated with nervousness or insincerity (Elaad & Gonen-Gal, 2022, pp. 2, 13), AI deception is smooth, continuous, and affectless.

This psychological vulnerability is worsened by confirmation bias, which refers to the tendency to seek or interpret information in ways that reinforce pre-existing beliefs (Nickerson, 1998, p. 175). This cognitive shortcut can lead people to build a narrative that fits their expectations. As Nickerson (1998) notes, if someone strongly believes they are a 'target of other people's ill will or aggression', they will likely be able to fit many otherwise ambiguous or unaccounted-for incidents into that pessimistic view, thereby confirming their own sense of being wronged (p. 182). Mediators can also (knowingly or unknowingly) find themselves giving more weight to evidence that fits the emotional flow of the session. While mediators do not formally evaluate evidence to determine its probative value as a judge would (Stempel, 1997), they are nonetheless influenced by the information presented. Their role in managing the dialogue means that evi-

dence which seems to confirm a developing narrative can subtly guide their interventions and determine the direction of the mediation. For instance, a mediator hearing a family succession dispute may have a pre-existing narrative in mind based on initial pleadings. A deepfake video appearing to confirm this narrative would be readily accepted due to confirmation bias, derailing the search for a more nuanced truth. This creates a loop in which fake content not only deceives but also reinforces existing biases. As Yuval Noah Harari (2019) warned in *21 Lessons for the 21st Century*, 'in a world deluged by irrelevant information, clarity is power. But clarity can be faked'. In mediation, clarity is derived from patient dialogue and emotional intelligence (Isabel, 2025). Now, clarity can be manufactured. It is clarity without truth.

This makes mediation, a process that is usually filled with emotions, particularly vulnerable (Kalter *et al.*, 2021). Mediators are taught and encouraged to listen carefully, observe nonverbal cues, and detect dissonance between what is said and what is meant (Fisher, Ury & Patton, 1991, p. 67). But these skills possessed by humans are somewhat useless when facing content that has been designed to appear authentic. What does 'body language' mean in a video that never happened? What does a voice reveal when it has been altered from a myriad of syllables fished from the internet? This is an epistemological challenge that mediators never dreamt of facing.

III. SYNTHETIC EVIDENCE IN MEDIATION AND ITS PROCEDURAL AND ETHICAL IMPLICATIONS

The introduction of synthetic evidence into mediation shifts the inquiry from psychological concerns to procedural and oversight matters. This has placed the mediation process, which is built on good-faith dialogue, on the front lines of a new kind of conflict: one in which fabricated narratives can poison the well of trust between the parties (Kakulungulu 2025). This begs the question: if a settlement is reached based on fabricated informa-

tion, what will protect the integrity of the process and the autonomy of the parties, given the mediator's non-adjudicative role?

Consider the Carbonaro Effect, popularized by Michael Carbonaro's television series, which demonstrated a psychological phenomenon in which people are led to believe that staged magic tricks are real because they trust their own eyes (Koehler, 2019). Today, the magician is an algorithm designed to manipulate human perception at a level of sophistication that exceeds human ability to detect it (Chong, 2025). A Massachusetts Institute of Technology (MIT) study revealed one significant vulnerability: when faced with subtle, high-quality fakes, one in four people failed to identify the altered image, even though they knew a manipulation was present (Groh et al., 2021, p. 43). If human perception can be so easily manipulated, what hope does a process built on authentic dialogue have? This question is primarily one of procedural integrity and design, not of professional liability.

Traditionally, the law has been respectful of mediators, trusting them to maintain standards of neutrality, voluntariness, and good faith. Under the Kenyan CAM Rules, mediation is defined as an 'informal and non-adversarial process' where the mediator's role is to 'encourage and facilitate' resolution, not to act as an adjudicator (Civil Procedure (Court-Annexed Mediation) Rules, 2022, r. 2). However, as the information parties bring to the table becomes more vulnerable to tampering, the line between facilitating dialogue and ensuring a fair process starts to become fuzzy.

While a deepfake has officially derailed no Kenyan mediation, the increasing digitalization of evidence and communication makes such a scenario plausible. Kenyan courts and ADR forums now routinely handle digital evidence such as WhatsApp messages, emails, and audio recordings (Koyoo, 2023, p. 86). The National ADR Policy itself recognizes the need to leverage Information and Communication Technology (ICT), which implicitly acknowledges the growing presence of digital materials in dispute resolution (Sessional Paper No. 4 of 2024 on the National

ADR Policy, p. 45). These digital challenges bring both opportunity and risk.

Across the world, these similar challenges have begun to flow into litigation and arbitration (Burgstaller & Macpherson, 2021; Amatika-Omondi, 2022, pp. 145-186). While it is crucial to avoid a doctrinal misfit by conflating mediation with adversarial proceedings, cases from formal adjudication provide the clearest warnings about the potential for AI to pollute dispute-resolution processes. For example, in the case of *Harber v Commissioners for His Majesty's Revenue and Customs* [2023] UKFTT 1007 (TC), it was found that cases cited by the litigant were not genuine judgements but had been created by an AI system such as ChatGPT. This was the first of its kind in the UK (Bowler, 2023). The Tribunal concluded that the cases cited by the appellant did not exist and had been fabricated to support her arguments. Although this specific case concerned textual fabrication and occurred within a formal tribunal with powers of scrutiny that mediators lack. It showed the ease with which digitally generated falsehoods can enter dispute-resolution forums, including those in Kenya.

Perhaps the most notable Kenyan case study on the dangers of AI is the arrest over an AI-generated image of the President (Daily Nation, 2025). In early 2025, Benson Malova Ashiko was arrested for allegedly posting an AI-generated image depicting a fabricated funeral of the President of the Republic of Kenya on his X (formerly Twitter) account (Daily Nation, 2025). This incident, though occurring outside a dispute-resolution setting, vividly illustrates how easily synthetic media can be created and disseminated in Kenya, and how the legal framework is unprepared to address such fabrications. The Data Protection Act, for instance, focuses on the misuse of genuine personal data, not the creation of synthetic data that impersonates individuals (Kiilu & Kiendi, 2022, p. 11). Indeed, the Act primarily addresses issues such as lawful processing, consent, and data retention, leaving a significant regulatory vacuum regarding novel forms of data

manipulation, such as deepfakes (The Data Protection (General) Regulations, 2021, regs 19, 22). This legal lacuna leaves courts and mediators alike in a precarious position.

While no Kenyan court has yet adjudicated a mediated settlement arising from synthetic evidence, recent jurisprudence on setting aside mediated agreements for procedural failures provides a strong analogical foundation, establishing that the integrity of the mediation process itself is a justiciable issue (*Kibosia & 11 others v Chebelieni & another*, [2024], para 14). For instance, in *Kibosia & 11 others v Chebelieni & another* [2024] KECA 1269 (KLR), the Court of Appeal invalidated a settlement not based on the substance of the property distribution, but due to a 'fundamental mistake' in the process, the failure of all parties to sign the agreement (Kibosia & 11 others v Chebelieni & another, [2024], para 28). The court held that the lack of signatures from all 24 beneficiaries meant there was no consensus, and therefore, 'no proper mediation settlement agreement for the court to adopt as a consent judgment' (*Kibosia & 11 others v Chebelieni & another*, [2024], para 28). This precedent is critical, as it suggests that a settlement procured through materially deceptive synthetic evidence could similarly be viewed as resulting from a fundamental procedural error, since one party's consent would be based on a fabricated reality, thus negating genuine consensus and rendering the agreement voidable.

Similarly, the case of *Tiego & another v Mahagwa & another* [2024] KEELC 6182 (KLR), while dismissing the application as premature, affirmed the procedural pathway for such challenges (*Tiego & another v Mahagwa & another* [2024], para 16). The court explicitly referenced Rule 39(3) of the CAM Rules, which allows a party to set aside a mediated decree on grounds including 'misconduct, fraud or fundamental mistake by the mediator' or 'fraud, collusion or misrepresentation by any party' (*Tiego & another v Mahagwa & another* [2024], para 14). This confirms that a legal mechanism already exists within Kenya's CAM framework through which a party victimized by a deepfake could seek

redress, arguing that their agreement was procured by fraud or a fundamental mistake of fact that materially affected their decision to settle (*Tiego & another v Mahagwa & another* [2024], para 14).

What complicates the situation is the power imbalance and access differences. In many disputes, one party may have greater financial resources or be more technologically savvy, enabling them to produce or commission convincing deepfakes. In contrast, the other party lacks the means to challenge them (Nyaga, Ondego & Joel, 2023, p. 9). It is the duty of the parties to act in good faith during the mediation process; however, this duty is often aspirational rather than enforceable, leaving a procedural gap that can be exploited (Nyaga, Ondego & Joel, 2023, p. 5). In such situations, if the process is influenced by concocted evidence, it might lead to an unjust outcome and procedural unfairness. Since mediation is party-driven and the mediator does not make decisions, the harm occurs when fabricated evidence distorts a party's perception of reality, leading them to agree to a settlement they otherwise would not have (Stempel, 1997, p. 958). This not only undermines the individual outcome but also erodes public confidence in mediation, a system the Kenyan Judiciary has heavily invested in to improve access to justice and reduce case backlogs (Judiciary of Kenya, 2023, p. 108).

The answer lies not in transforming mediators into digital forensic experts, but in rethinking the support systems around them and their role as active guardians of a fair process. It means professional standards must evolve to include digital literacy and protocols for addressing suspicious material. For example, amending mediation rules to allow for, upon the parties' agreement, the referral of a contentious digital file to a neutral third-party expert for verification. This would not involve the mediator making an evidentiary ruling, but rather facilitating a solution to a procedural roadblock that threatens the integrity of the dialogue—a role consistent with a mediator's function to 'guide the mediation process' (Civil Procedure (Court-Annexed

Mediation) Rules, 2022, r 22). The process of caucusing (holding private meetings with each party) offers the mediator an opportunity to sensitively probe concerns about evidence without escalating conflict in a joint session (Nyaga, Ondego & Joel, 2023, p. 3). In conclusion, addressing synthetic media goes beyond simple culpability; it involves a responsibility for the integrity of future settlements.

IV. TECHNOLOGICAL SOLUTIONS AND THEIR SHORTCOMINGS

Faced with the threat of AI-generated deception, a logical response is to turn to technology for the solution (Park *et al.*, 2024, p. 11). AI and technological tools can be deployed to verify the authenticity of evidence in mediation; however, they can be fallible and inaccessible. The argument is not that authentication tools are incompatible with mediation's objectives; rather, the challenge lies in integrating them without fundamentally altering the process. Verifying the authenticity of evidence is a procedural step that can, in fact, support the ultimate goal of understanding party narratives and repairing relationships (Coben, Fox & Love, 2018, pp. 14-15, 18). The danger arises if the process becomes dominated by technical disputes over authenticity, thereby sidelining the human-centric dialogue that mediation is meant to foster (Gillett, 2025). The objectives of mediation go beyond identifying the 'true' documents or the 'correct' video clip. It is about understanding each party's arguments, repairing broken relationships, and restoring the dignity of the parties (Golann, 2002).

If AI can create convincing fakes, perhaps a more sophisticated AI can detect them (Park *et al.*, 2024, p. 11). This line of reasoning has given rise to a burgeoning field of media forensics, where algorithms are trained to identify the subtle, often invisible artifacts left by deepfake creation processes (Verdoliva, 2020). These detection tools work by analyzing inconsistencies that the

human eye might miss, such as unnatural blinking patterns, reflections in eyeglasses that are distorted or absent, blurring at the edges of a face, or anomalies in pixel noise and compression (Park *et al*., 2024, p. 12). For audio deepfakes, similar tools can analyze spectrograms to detect non-human frequencies or unnatural transitions between phonemes (Pham *et al*., 2024). In theory, these AI-powered verification tools could serve as a crucial line of defense in mediation, offering a seemingly objective way to flag suspicious evidence before it can poison the dialogue.

However, this technological solution has its own pitfalls, creating a high-stakes 'arms race' between forgery and detection (Laurier, 2024). As deepfake generation models become more advanced, they learn from the techniques used to detect them, producing forgeries that are progressively harder to identify (Laurier, 2024). This creates a scenario where any detection tool may have a very short shelf life before it is rendered obsolete by the next generation of generative AI (Groh *et al*., 2021). Furthermore, over-reliance on these tools introduces new challenges to the mediation process.

First, the issues of accessibility and cost. In Kenya, access to and training on such sophisticated forensic tools are not standard for accredited mediators (Barasa, 2025, p. 6). The National ADR Policy envisions leveraging ICT to enhance accessibility. Still, it does not budget for or mandate the provision of advanced forensic software to mediators (Sessional Paper No. 4 of 2024 on the National ADR Policy, p. 45). Who would bear the cost of an AI detection tool or a neutral expert in a CAM proceeding? The parties, who are often in mediation precisely to avoid the high costs of litigation? Or the Judiciary, whose budget is already strained, facing a fifty-one percent funding gap in the 2022/2023 fiscal year (Judiciary of Kenya, 2023, p. 260)? The Sessional Paper on the National ADR Policy advocates for leveraging ICT but remains silent on the funding mechanisms for such specialized tools (Sessional Paper No. 4 of 2024 on the National ADR Policy, p. 44). Without a clear cost model, these technological solutions

remain theoretical for the vast majority of Kenyan mediations. Sophisticated forensic tools may be financially out of reach for one party (Amatika-Omondi, 2022, p. 164), creating a significant power imbalance where the more resourced party can both produce and challenge digital evidence more effectively. This risks turning mediation, intended as an accessible alternative to court, into a battle of competing experts and expensive software, undermining its core purpose.

Second, and perhaps more insidiously, is the problem known as the 'liar's dividend' (Citron & Chesney, 2019, pp. 1758, 1785). As public awareness of deepfakes grows, a malicious actor can dismiss genuine, incriminating evidence by simply claiming it is a deepfake. In mediation, where there is no judge to make a definitive ruling on authenticity, such a claim can weaponize doubt and sow enough confusion to derail the entire process. A party could use the mere possibility of a deepfake to cast aspersions on legitimate video or audio recordings, making it impossible for the parties to agree on a common set of facts. This erodes trust not only in fake content but in all digital evidence, potentially paralyzing the dialogue (Kakulungulu 2025). Therefore, while AI detection tools are a necessary part of the response, they are not a silver bullet. They must be viewed as one component of a broader strategy that also strengthens human discernment and procedural safeguards, acknowledging that code alone cannot restore a compromised sense of reality.

V. RETHINKING VIGILANT GUARDIANSHIP IN MEDIATION

If AI has taught humanity anything, it is that being human is not a design flaw (Montemayor *et al*., 2022). It is a design feature. For centuries, mediators have held fast to the ideal of neutrality, not as indifference, but as a delicate equilibrium: a posture of fairness, a listening ear equally inclined to both sides. It is a human art grounded in intuition, empathy, and the silent calibration of power and emotion in the room. However, in this

emerging architecture of digital dispute resolution, where voices and videos can be easily tampered with by a single click, that ideal is being redrawn. The future of mediation still calls for impartiality. Still, it will also require an intentional reckoning with bias, a reaffirmation of human limits, and a courageous embrace of the irreducible qualities that no code can replicate.

Neutrality, which was once based on professional decorum and codes of ethics, is now trapped in a labyrinth of intuition, cognition, and code. The same qualities that enabled a human mediator to connect and relate are also the areas where bias may reside. Cognitive science reminds one that no human decision-making process is free of bias (Banaji & Greenwald, 2013). Implicit biases, developed by race, gender, class, culture, and experience, operate below the level of consciousness (Banaji & Greenwald, 2013). They shape judgments in ways one hardly realizes. In Kenya, while the Code of Ethics for mediators under the MAC emphasizes impartiality and independence, it does not explicitly address the cognitive biases that synthetic media can exploit (Judiciary of Kenya, Mediation Accreditation Committee Code of Ethics).

The shift from a passive neutral to a 'vigilant guardian' does not require abandoning the core tenets of the MAC's Code of Ethics for Mediators, but rather reinterpreting them for the digital age. For instance, the duty of 'professional competence' under paragraph 12 must now be understood to include a baseline level of digital literacy (Mediation Accreditation Committee, 2022, p. 8). The duty to maintain the 'quality of the process' under paragraph 8 must encompass a responsibility to guide the parties through procedural challenges posed by suspicious digital evidence (Mediation Accreditation Committee, 2022, p. 6). This is not a departure from impartiality but an affirmation of it, as ensuring a procedurally fair dialogue is central to a mediator's role. However, no matter how carefully a law is crafted, none can exorcise the ghost of human prejudice.

This brings forth an uncomfortable enigma: human mediators are susceptible to cognitive biases that deepfakes can exploit; technological tools have their own limitations and can introduce new forms of inequality. Where then does that leave neutrality? The solution is not to strip mediation of its human element, but to evolve it (Stempel, 1997, pp. 975, 984). Mediation, at its very core, is not merely the resolution of disputes; it involves mending fractured relations (Bush & Folger, 2005). It is a process that helps parties face their trauma, acknowledge their circumstances, and restore their dignity, not by the dictates of the law but by the posture of the listener. This is where the law has to evolve, not by entirely abandoning the principle of neutrality, but by redefining it for this new era.

For ages, neutrality has always been defined in procedural terms: a neutral person avoids conflicts of interest, does not favor any party, and refrains from imposing outcomes (UNCITRAL, 2004). But today, in a world filled with digital deception, neutrality must mean more than passive detachment. It must mean active discernment. The Book of Proverbs warns that, 'The simple believe every word: but the prudent man looketh well to his going' (Proverbs 14:14, King James Bible, 2017). This is not just a call for cynicism but for critical neutrality: a model of impartiality that recognizes human intuition is fallible and requires the mediator to become a vigilant guardian of the process itself. The mediator is now forced to walk on eggshells: emotionally attuned yet procedurally vigilant, empathetic yet skeptical, human yet digitally literate.

A. Advancing the Codification of Digital Vigilance and Ethical Foresight

To address this intersection between technology and mediator neutrality, Kenya will need to adopt regulations that anticipate digital disruptions and protect the essential principles of impartial dispute resolution. This requires the implementation of new legislation(s), professional standard(s), and encouraging capacity-building.

First, the Parliament must revise and expand the Civil Procedure Act (Court-Annexed Mediation) Rules, 2022, to expressly recognize the risks posed by manipulated digital evidence. Even though this legislation establishes a strong framework for the procedure, it is currently silent on the verification of digital evidence, a significant gap given that the National ADR Policy acknowledges the sector suffers from 'technological gaps' and requires unified standards (Sessional Paper No. 4 of 2024 on the National ADR Policy, p. 26). To rectify this, the Rules should include a duty of procedural vigilance requiring mediators to take reasonable steps when seemingly dispositive digital documents, videos, or AI-generated materials are introduced. This duty would not need mediators to become forensic experts, but rather to know when to pause the process and facilitate a discussion between the parties on how to handle a contested piece of digital evidence, including the option to engage a neutral expert.

Specifically, in consultation with the MAC, the Rules Committee acting under its statutory mandate in Section 59A(4)(d) of the Civil Procedure Act, should consider introducing a new provision, perhaps as Rule 22A, stating: 'Where a party raises a good-faith challenge to the authenticity of digital evidence presented during mediation, the mediator shall facilitate a discussion on procedural options, which may include:

i). proceeding without the contested evidence;

ii). adjourning to allow parties to obtain authenticating information; or

iii). with the express written consent of both parties, referring the evidence to a neutral expert from a roster maintained by the Mediation Accreditation Committee for a non-binding authenticity assessment'.

Second, the MAC should make good use of its authority to set standards (Judiciary Kenya, 2017) by formulating and disseminating Practice Directions and Guidelines for Digital Evidence in mediation. These guidelines should be practical and

accessible, providing simple red flags for identifying potential forgeries and a clear, step-by-step protocol for mediators to follow. For instance, a protocol could involve:

i). pausing the mediation;

ii). caucusing with each party to understand their position on the evidence;

iii). exploring options with the parties, such as agreeing to disregard the proof or jointly commissioning a neutral expert; and

iv). documenting the procedural challenge and the agreed way forward in the mediator's report (Nyaga, Ondego, & Joel, 2023, p. 3).

Although the Evidence Act of Kenya (CAP. 80) already accommodates electronic evidence under Sections 106A and 106B, mediation still requires its own ancillary rules to prevent informal processes from being exploited through technical loopholes. The most important question is that mediation's informality, its core strength, is also its primary vulnerability to deepfake (Karanja, 2025). Therefore, these ancillary guidelines are not about importing litigation's rigidity, but about creating a bespoke, light-touch framework that addresses a vulnerability unique to mediation's informal structure.

Third, drawing on instruments such as the European Commission for the Efficiency of Justice (CEPEJ) *European Ethical Charter on the use of Artificial Intelligence (AI) in Judicial Systems* (2018), the MAC should develop principles for the use of AI tools in mediation support. While this paper focuses on the threat of deepfakes, any framework must also govern the use of technology as a solution. The CEPEJ Charter emphasizes core principles such as respect for fundamental rights, non-discrimination, quality and security, transparency, and most importantly, the principle of being 'under user control' (CEPEJ, 2018, p. 12). Adopting similar principles would ensure that any AI tool used for evidence verification in mediation must be understand-

able to the parties and subject to the final oversight of the human mediator, preventing the process from being dictated by a 'black box' algorithm. This safeguard would ensure that decisions are not shaped or distorted by unverified outputs, thus maintaining both integrity and neutrality.

Finally, the role of professional liability must be clarified. While mediators are generally exempted from civil liability under the Mediation Rules of 2022 (provided they acted in good faith), this immunity ought not to extend to reckless disregard or willful blindness of forged evidence. There is a fragile line between trust and professional negligence. Just as lawyers are bound by duties of competence under the Advocates (Practice) Rules, so should mediators also be held to account where their failure to interrogate obviously suspicious digital material causes demonstrable harm or injustice. To attain this objective, the Law Society of Kenya (LSK), together with the MAC, and other ADR institutions like the Nairobi Centre for International Arbitration (NCIA), which has been identified as the national oversight body in the National ADR Policy (Sessional Paper No. 4 of 2024, p. 42), should co-develop a Code of Digital Ethics for Mediators.

Most importantly, this must be supported by mandatory Continuous Professional Development (CPD) modules on digital evidence and ethics, a capacity-building measure that was identified as critical in the National ADR Policy (Sessional Paper No. 4 of 2024 on the National ADR Policy, p. 44). These mandatory CPD modules for mediators should include:

i). 'Digital Evidence 101': Understanding metadata, hashing, and common forms of digital manipulations;

ii). Ethical Practice in the Digital Age': Navigating the 'liar's dividend' and maintaining neutrality; and

iii). 'Procedural Protocols': A step-by-step guide for handling evidentiary challenges without assuming an adjudicative role.

This implements the capacity-building objective of the National ADR Policy (Sessional Paper No. 4 of 2024, p. 44). Likewise, it would equip mediators with the necessary skills to identify potential digital manipulations and manage the procedural challenges that arise, fulfilling the policy's goal of enhancing the quality of ADR services in Kenya (Sessional Paper No. 4 of 2024 on the National ADR Policy, p. 42). To operationalize the use of neutral experts, the MAC should establish and maintain a roster of vetted digital forensics experts. If parties agree to use an expert, the process should be governed by a standard tripartite agreement between the parties, the expert, and the mediator, outlining the expert's limited mandate (authenticity verification only), confidentiality obligations, and cost-sharing arrangements.

VI. CONCLUSION

Trust is the foundation of mediation. It is trust that enables parties to feel free to speak without constraints, to confess in privacy, to concede publicly, and to maintain faith in the process when results are less than ideal. Yet in a world with a lot of synthetic 'truths', AI-generated deepfakes, and distorted digital documents, trust cannot remain stagnant. It needs to evolve. The famous Swahili proverb 'mwenye macho haambiwi tazama' (he who has eyes cannot be told to see) can no longer prevail (Kitula, 2009). Today, in the age of deepfakes and voice clones, even what is seen and heard must be interrogated. Therefore, mediators who were once just neutral vessels must become the custodians of discernment. This evolution does not mean abandoning the human element. On the contrary, it necessitates promoting human judgment to counter the challenges posed by machine-made manipulation.

The law, too, must rise to this moment. Laws governing mediation can no longer treat digital distortion as an outlier. They need to enshrine responsibility, strengthen procedural protec-

tions, and implement training processes that will prepare mediators not just to listen sympathetically but to think with acuity. Neutrality in the near future will not just mean lack of bias, but a conscious engagement with it: a neutrality that sees, discerns, questions, and holds itself accountable. This is more a matter of law than it is a theoretical shift. When courts start to review mediated settlements, especially under frameworks like the Singapore Convention on Mediation, they will have to test whether neutrality was not only claimed but also exercised in technologically demanding environments.

To mediate in the era of technology is to straddle two worlds: one of emotions, trauma, and human frailty; and another of code, data, and computational logic. The future mediator has to be fluent in both. They must be, as Proverbs recommends, both 'simple' in enthusiasm and 'prudent' in vigilance. For only at their intersection can mediation preserve its most precious offering: justice that is not only delivered but believed.

REFERENCES

Amatika-Omondi, F., The Regulation of Deepfakes in Kenya (2022) 2(1) *Journal of Intellectual Property and Information Technology Law (JIPIT)* 145.

Arendt, H. (1973). *The origins of totalitarianism* (New ed.). Harcourt Brace Jovanovich.

Banaji, M. R., & Greenwald, A. G. (2013). *Blindspot: Hidden biases of good people*. Delacorte Press.

Barasa, G.M, 'Mediation Ecosystem in Kenya: Current Trends and Future Prospects' [2025] Nairobi Centre for International Arbitration.

Bowler, Z. (2023, December 19). One of the first domestic judicial decisions highlighting the risks of relying on artificial intelligence (AI) in case preparation has been handed down by the First-Tier Tribunal (FTT) Tax Chamber. *Thompson Snell and Passmore*. https://www.ts-p.co.uk/insights/landmark-ruling-first-tier-tax-tribunal-warns-against-ai-reliance-in-case-preparation/

Boyd, D. (2017, February 23). Danah Boyd on the spread of conspiracies and hate online. *PBS NewsHour*. https://www.pbs.org/video/danah-boyd-how-digital-media-can-spread-false-information-en/

Burgstaller, M., & Macpherson, S. (2021). Deepfakes in international arbitration: How should tribunals treat video evidence and allegations of technological tampering? *The Journal of World Investment & Trade*, 22(1).

Bush, R. A., & Folger, J. P. (2005). *The promise of mediation: The transformative approach to conflict*. Jossey-Bass.Chong, N. (2025, January 1). The rise of the deceptive machines: When AI learns to lie. *United Nations University Computing Centre*. https://c3.unu.edu/blog/the-rise-of-the-deceptive-machines-when-ai-learns-to-lie

Chowdhury, S. M. A. K., & Lubna, J. I. (2020). Review on Deep Fake: A looming technological threat. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT).*

Citron, D. K., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy and national security. *California Law Review*, 107, 1753.

Civil Procedure Act, The Civil Procedure (Court-Annexed Mediation) Rules, 2022.

Cloke, K. (2017). *Mediation Training Manual*. Center for Dispute Resolution.

Coben, J., Fox, K., & Love, L. P. (2018). A Conversation about facts, stories, and truth in mediation. *Dispute Resolution Magazine*, 14(4), 14.

Elaad, E., & Gonen-Gal, Y. (2022). Face-to-Face Lying: Gender and Motivation to Deceive. *Frontiers in Psychology*, 13.

European Commission for the Efficiency of Justice (CEPEJ). (2018). *European Ethical Charter on the use of Artificial Intelligence (AI) in judicial systems and their environment*.

Evidence Act, Cap. 80, Laws of Kenya.

Fisher, R., Ury, W., & Patton, B. (1991). *Getting to yes: Negotiating agreement without giving in*. Houghton Mifflin.

Gaffney, I. (2022). Impartiality and neutrality in mediation. *The Journal of Mediation and Applied Conflict Analysis*, 8(1).

Gilbert, D. T., Malone, P. S., & Krull, D. S. (1990). Believing the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59(4), 601-613.

Gillett, E. (2025, June 1). The Double-Edged Sword: AI in Mediation and What We Need to Watch For. *King County Bar Association Bar Bulletin*. https://www.kcba.org/?pg=News-Bar-Bulletin&blAction=showEntry&blogEntry=126105

Golann, D. (2002). Is legal mediation a process of repair – or separation? An empirical study, and its implications. *Harvard Negotiation Law Review*, 7, 301–340.

Graber-Mitchell, N. (2021). Artificial intelligence: Deepfakes as speech. *Intersect: The Stanford Journal of Science, Technology, and Society*, 14(3).

Groh, M., Epstein, Z., Obradovich, N., Cebrian, M., & Rahwan, I. (2021). Technologies for manipulating and faking online media may outpace people's ability to tell the difference. *Communications of the ACM*, 64(10).

Hao, J. (2024). Advances in deepfake generation and detection technologies: Challenges and opportunities. *Transactions on Computer Science and Intelligent Systems Research*, 6.

Harari, Y. N. (2019). *21 lessons for the 21st century*. Random House.

Harber v Commissioners for His Majesty's Revenue and Customs [2023] UKFTT 1007 (TC).

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5), 1191-1206.

Hunt, K., & Chittka, L. (2014). False memory susceptibility is correlated with categorisation ability in humans. *F1000Research*, 3(154).

Isabel, D. (2025). The emotional core of conflict resolution: Integrating emotional intelligence into mediation. *Journal of Appropriate Dispute Resolution (ADR) & Sustainability*, 3(1).

Judiciary of Kenya. (2022). *Mediation Accreditation Committee: Code of Ethics for Mediators*.

Judiciary of Kenya. (2023). *State of the Judiciary and Administration of Justice Annual Report 2022-2023*.

JusCorpus. (2023, July 20). Robert Nozick's entitlement theory of justice: A critical analysis. *JusCorpus*. https://www.juscorpus.com/robert-nozicks-entitlement-theory-of-justice-a-critical-analysis/

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kakulungu, S. (2025). Deception in legal proceedings: Ethical implications. *Eurasian Experiment Journal and Management*, 7(1).

Kalter, M., Bollen, K., Euwema, M., & Verbeke, A-L. (2021). A matter of feelings: Mediators' perceptions of emotions in hierarchical workplace conflicts. *Frontiers in Psychology*, 12.

Karanja, G. W. (2025). Law Over Borders Comparative Guide: Mediation Law Guide - Kenya. *The Global Legal Post*, May 2025.

Keyes, R. (2004). *The post-truth era: Dishonesty and deception in contemporary life*. St Martin's Press.

Kiilu, J., & Kiendi, L. (2022). Consent for Processing Children's Personal Data in Kenya – New Dawn in Kenya Privacy Laws. *Law Society of Kenya Journal*, *18*, 7.

King James Bible. (2017). *King James Version*. Cambridge University Press.

Kitula, G. (1989). *Kamusi ya methali za Kiswahili*. East African Educational Publishing.

Koehler, S. (2019, July 30). The carbonaro effect: 14 facts only true pranksters will know about the show. *Looper*. https://www.looper.com/160280/the-untold-truth-of-the-carbonaro-effect/

Koyoo, N. (2023). Pathway to Digital Justice and Innovative Mechanisms: An analysis of amicable dispute settlement in Kenya. *University of Nairobi Law Journal*, *10*(2), 84.

Laurier, L., Giulietta, A., Octavia, A., & Cleti, M. (2024). The cat and mouse game: The ongoing arms race between diffusion models and detection methods. *arXiv preprint*.

Maute, J. L. (1990). Mediator accountability: Responding to fairness concerns. *Journal of Dispute Resolution*, 1990(2), 347-367.

Montemayor, C., Halpern, J., & Fairweather, A. (2022). In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. *AI & Society*, 37, 1-9.

Munguti, R. (2025, January 28). Another social media user in court for posting fake news about Ruto. *Daily Nation*. https://nation.africa/kenya/counties/nairobi/another-social-media-user-in-court-for-posting-fake-news-about-ruto--4903804

Najdowski, C. & Stevenson, M. (2018). *Criminal juries in the 21st century: Contemporary issues, psychological science and the law*. Oxford University Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.

Nyaga, B. M., Ondego, J. C., & Joel, M. (2023). Mediation and Data Protection Law in Kenya: Appraising ADR for Optimal Access to Justice under the DPA 2019. *SSRN Electronic Journal*.

Office of the Attorney General and Department of Justice. (2024). *Sessional Paper No. 4 of 2024 on the National Alternative Dispute Resolution Policy*.

Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. *Patterns*, 5(5).

Pham, L., Lam, P., Nguyen, T., Nguyen, H., & Schindler, A. (2024). Deepfakes audio detection using spectrogram-based feature and ensemble of deep learning models. *arXiv preprint*.

Prigoff, M. (1988). Toward candor or chaos: The case of confidentiality in mediation. *Seton Hall Legislative Journal*, 12(1), 1.

Qureshi, J., & Khan, S. (2024). Deciphering deception- the impact of AI deepfakes on human cognition and emotion. *Journal of Advances in Artificial Intelligence*, 2(1).

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8(3), 338-342.

Rutenberg, I., Kiptinness, S., & Sugow, A. (2021). Admission of Electronic Evidence: Contradictions in the Kenyan Evidence Act. *Digital Evidence and Electronic Signature Law Review*, *18*, 35.

Sandoval, M.-P., de Almeida Vau, M., Solaas, J., & Rodrigues, L. (2024). Threat of deepfakes to the criminal justice system: A systematic review. *Crime Science*, 13(1), 41.

Schulz, J. (2023). Mediator liability 23 years later: The 'Three C's' of case law, codes and custom. *Ottawa Law Review*, 55(1).

Stempel, J. W. (1997). Beyond formalism and false dichotomies: The need for institutionalizing a flexible concept of the mediator's role. *Florida State University Law Review*, 24.

United Nations Commission on International Trade Law. (2004). *UNCITRAL model law on international commercial conciliation with guide to enactment 2002*. United Nations.

Verdoliva, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 910-932.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11).

Williams, K., & Hinshaw, A. (2018). Outbursts: An evolutionary approach to emotions in the mediation context. *Negotiation Journal*, 34(3), 263-281.