Beyond A Reasonable Doubt? Audiovisual Evidence, AI Manipulation, Deepfakes, and the Law

***Abstract*—THE CAPTURE is a mystery thriller series, that completed its second season on Peacock and BBC One. The British television drama revolves around the alteration of direct audiovisual evidence on the command of a special unit that believes there is enough circumstantial evidence to either convict or acquit an individual of a felony. Based on the plot of the television series, this paper explores the potential for a variety of AI-enabled applications to be used in the course of criminal proceedings. The implications of evidence tampering are considered through AI manipulation toward the realization that deepfake evidence may well be admitted in court dependent on the human decision-maker. Will the future demand the interpolation of visual evidence for high profile criminal cases, and what does the existence of Generative AI and deepfakes mean for the forensic analysis of audiovisual evidence? After contemplating the socio-technical plausibility of the central premise of THE CAPTURE, this paper then turns to its legal implications. Drawing on examples from U.S. and Australian legal frameworks, the paper considers the consequences of AI-corrected, augmented or generated audiovisual evidence on three facets of natural justice: (1) the presumption of innocence; (2) the fair trial; and (3) lawyers' ethical duties of competence and to the administration of justice. The key takeaways of the paper are that: (1) deepfake evidence will continue to proliferate; (2) that the law will need to address both the substantive and procedural impacts of such evidence, and (3) that the legal profession must continue to educate its lawyers and practitioners, and associated stakeholders, of the nature, uses and risks posed by deepfake audiovisual artefacts to maintain public trust in the legal system.**

***Index Terms*—Audiovisual recording technology, GenAI, deepfakes, evidence, law enforcement, criminal investigations, forensics, auditing tools, law, lawyers, courts, fair trial.**

I. Introduction

IN A WORLD that has placed all its hopes on a digital economy, our bodies and indeed our faces have become interlinked with our experience of autonomy and our understanding of the essence of "who we are". Advances in the analysis and use of biometrics, however, present a challenge to this experience. Everything from the automatic identification of humans, to the electronic payment of goods and services, to the ability to travel internationally using an e-passport, and even share personal stories online through social media, requires our identity to be verified. Where once it might have been possible to live "off-the-grid", without an e-mail address or even a mobile number, one cannot live without biometrics. This is because biometrics – our biological, physical and behavioral characteristics – are not about us, they *are* us. Everyone has DNA. Everyone has a unique way of moving through the world. And everyone has a head on their shoulders, despite that the face may present as markedly different based on many factors: ethnicity, fashion, (plastic) surgery, natural or human-made markings (e.g., tattoos or other markings), disability, weight loss/gain, and more. Increasingly, in the name of security, our biometrics are being used in a tokenized manner, devoid of any semblance to their natural and organic states.

A. Biometrics as a Security Mechanism in a Digital Economy

Although biometrics are not new, security threats have meant that new methods of registration and enrolment using a variety of identification techniques have been designed, developed, piloted, and implemented in large scale settings, predominantly to date, by government services that are linked to government payments for vulnerable citizens, inclusive of the unhoused, unemployed, people living with disabilities, and pensioners. With the modern emphasis, as reflected in law [1], on making systems more robust and secure, public and private organizations, institutions and agencies have flocked to more robust forms of complex algorithms for identification, data encryption, and identity tokens that are unique, permanent and universal in opposition to easily replicatable tokens that may be counterfeit and offer little assurance against unauthorized access, collection and usage.

This does not mean that the collection and storage of biometric data for age assurance and identity verification is not subject to hackers who are able to penetrate the physical boundaries of an organizational entity via duplication, duping, masquerading, and impersonation. These are all security attacks that fall under the category of social engineering with the aid of technical attacks like phishing. Today, two-factor authentication strategies are recommended to guard against the possible overcoming of biometric defenses, by the acquisition of a facial scan as a first step, and then ensuring a second line of defense using a network-based handheld token or password that the user is in possession of via a personal device such as a smartphone. By disallowing and making it more difficult for hackers to access and copy identification numbers and passwords alone, biometrics has offered a way forward in the electronic identification of citizenry (e.g., Aadhaar multimodal biometric system in India of an estimated 1.27 billion people, with 93% saturation of the total market [2]).

Digital Object Identifier 10.1109/TTS.2024.3427816

B. Insecurity and the Changing Technological Landscape

For anyone who has studied emerging technologies, it becomes quickly apparent that the more we try to secure our digital systems, the more they – and, in turn, we - are subject to insecurities. That is, the more we tighten the ways in which we believe we will garner even greater security using advanced techniques, the more sensitive the individuals' relinquished information becomes, which are paradoxically subject to data breaches, and the compromising of once seemingly foolproof solutions. One hundred percent security is a misnomer. It cannot exist. The outside forces at play will impact a technology's security defenses in ways that were never previously considered. Biometrics, first manual then automated, may have considered typical "presentation attacks" and other potentialities, but morphing attacks, and deepfakes, may well not have been on the agenda [3]. Today, developments in Generative Artificial Intelligence (GenAI) catapult these attacks into real vulnerabilities. It is in this context that this paper looks to fiction film to consider the possibilities and challenges that we face, not only with respect to biometric technologies, but also with respect to what might be considered ethical, legal, plausible, and possible potentialities in the uses of our appearance, mannerisms, spoken language, dialect, and spoken word, behaviors or even likeness.

C. Outline

In Section II we briefly explore the methodology of this paper, in terms of the single case study used and analyzed, to elucidate the role of science fiction in examining the socio-legal implications of emerging technologies. In Section III we use THE CAPTURE series to explore the plot focused on deepfakes and the use of AI-enabled audiovisual capabilities which in essence relies on biometric identification and verification. THE CAPTURE is then compared against THE FINAL CUT where the act of deletion of incriminating evidence found in video surveillance footage is juxtaposed against the notion of "correction" applied in THE CAPTURE. In Section IV we ponder on the use of audiovisual evidence as direct evidence in a court of law and emphasize the need for a chain of custody that can forensically determine whether an audiovisual record has been tampered with (i.e., "cut" or "corrected"). In Section V we present a discussion on the legal implications of audiovisual innovation in the context of deepfakes. It is here, we consider the consequences of GenAI on key hallmarks of natural justice, including: the presumption of innocence, the fair trial, and lawyers' ethical duties. In this section examples are drawn from United States (US) and Australian legal frameworks, as approaches indicative of common law jurisdictions.

II. Methodology

A. Studying Film to Consider Socio-Technical Potentials

Stories and storytelling can provide powerful lenses through which to examine the implications of emerging technologies. In this paper, we draw attention to two interrelated functions served by narratives present in science-fiction film. The first, relates to the importance of scenario-planning in the development of emerging technologies. As the diverse impacts of technologies outpace the regulation and governance of new product and process innovations, it is through imagination that we are able to capture the social, ethical and, indeed, legal dilemmas of our time. For this reason, narrative has become a mechanism by which to investigate transdisciplinary responses to the socio-ethical implications of technology [4]. Adopting an explorative approach, the hope of stories is to gain an appreciation of those things that may drive societal forces toward the construction of artefacts for living and working, and how these artefacts might affect humans and the environment at large. What is the purpose of the tools we are creating? When stories take on form through film, they are a powerful tool to ponder the future [5].

Scenario planning is a process that is embedded in the creation of new innovations [6]. Without the use of scenarios, we may be unable to anticipate or understand the (un)intended consequences, or potential risks, associated with emerging technologies. While ethics and other values are often bolted on as an afterthought to the creation of new technologies [7, p. 4], the growing complexities of digital ecosystems mandate that we - *from the outset* - speculate on what might be, so that we are better able to apply precautions or anticipatory governance approaches, and develop sustainable legal frameworks, that allow socio-technical systems to achieve their intended purpose [8]. With respect to deepfake technology, their application is not always harmful [9], and even biometrics can be useful in matching human faces, despite the exceptions.

The second function of science-fiction film relates to the appeal of the storytelling format to diverse audiences, beyond those engaged in the technology development process, or who are being prepared for the workforce. Citizens, through film, are given an entry point by which to engage in debate over the emergence of new technologies and how they might be (mis)used in everyday life. In this manner, a multi-stakeholder audience is reached through film that is agnostic to one-dimensional casting [103]. We all at once carry multiple roles in society, both as citizens and other job roles, such as developers, producers, public servants, volunteers, media and more. Film reaches everyone.

B. Single Case Study: THE CAPTURE Series as Stimuli

In the first episode of Season One of THE CAPTURE [10], closed circuit television (CCTV) footage emerges of Shaun Emery, one of the season's protagonists, attacking his barrister Hannah Roberts near a bus stop. Her body is later found, and when the last seen video surveillance recording of Roberts places her with Emery, who is a former British veteran who served in Afghanistan, things look grim for him. Emery, who had been accused of killing a soldier at point blank in Afghanistan, had denied the charges and won the murder case through a so-named "technical timing fault" in the CCTV footage captured in the war setting. But when seemingly undeniable video surveillance surfaces in the Roberts case, it leaves no way for Emery to raise a reasonable doubt. In this way, the first episode of THE CAPTURE calls into

question the adage that "the camera never lies" [11]. In fact, the prescience of this plot is that the adage no longer holds true in our digital age, despite the probative weight we attach to images and videos [12].

C. Qualitative Analysis Through a Socio-Legal Lens

Throughout the remainder of the paper, the single case study of THE CAPTURE is deployed as the main stimuli/intervention by which to qualitatively analyze the challenges that arise when we consider the role that new biometric technologies and corresponding AI-enabled audiovisual capabilities, new techniques, new forms of evidence, and new ways of adjudicating crimes may happen. In doing so, we are attempting to anticipate some of the problematic contexts that may ensue to assess the agility of the existing socio-legal infrastructure.

Tampering with audiovisual evidence is one thing, but the successful performance of real-time deepfakes can no longer be dismissed as implausible. There are now cases of successful deepfakes that have cost company's money, and tried and tested morphing attacks that have breached even the strongest biometric defenses, not to mention the ability for anyone to scrape the Web for the amassing of textual, image-based and audiovisual data (e.g., Clearview AI). Some of these cases began with highly sophisticated impersonation scams and then escalated with the use of deepfakes. For example, in 2020 one Argentinian gang leader stole USD 600,000 in just 10 days through a series of impersonation scams, and then in 2021 an employee impersonated the CEO of a Chilean mining company and stole USD 1.5 million [13].

But today, these scams incorporate believable deepfake live videos. For example, in February 2024, a finance worker in Hong Kong paid out USD 25 million after multiple "people" appeared on a deepfake video call, inclusive of someone masquerading as the company's U.K.-based chief financial officer [14]. In May in Inner Mongolia it was reported that a perpetrator used "face-swapping technology" to impersonate a friend of the victim where USD 622,000 was transferred, in the "belief that his friend needed to make a deposit during a bidding process" [15]. These new forms of technology-assisted impersonation hacks require various people to collaborate in the execution of a multiphase deepfake process, entailing a series of specializations. These include: (1) identifying the target victims, (2) creating the script using language known to the targets, (3) initiating the phishing email, (4) investigating the technology to be adopted and applied in real-time, (5) assigning actors to appear live and deliver the final dupe, (6) establishing a legitimate bank account used to transfer the stolen monies into and, finally, (7) developing an exit strategy.

GenAI companies, such as OpenAI are now using open-source intelligence (OSINT) to inform responses to prompt queries. The question remains, will these companies in the not-too-distant future become complicit to acts of biometric morphing attacks on people. This involves the use of uncontested algorithms, either in the form of Generative Adversarial Networks (GANs) or, most recently, Diffusion Models (DMs), applied to regulated environments for potential misuse of copyrighted materials, "filling the gaps" via hallucinations [16]. While these practices may not be expressly prohibited by law, most would call them out for being grossly unethical. The admonition of OpenAI's CTO in March 2024 that "publicly available data and licensed data" was used in SORA-generated video production was telling. The CTO proceeded to comment that she was unsure whether YouTube, Facebook or Instagram videos were used, when prompted by the *Wall Street Journal* reporter [17]. In the remainder of this article, we demonstrate these socio-technical tensions by studying the legalities emerging from the plot of film series THE CAPTURE.

III. Audiovisual Everywhere

A. Innovation in Video Surveillance

Video surveillance has greatly developed in the last 50 years. One of its earliest depictions in movies was in the original BATMAN (1966) starring Adam West, where the Joker, Penguin, Riddler, and Catwoman were featured on a single television monitor in the Gotham City Police Commissioner's office through video transmission signal [18]. Today, modern video surveillance of control rooms uses dedicated digital video recorders (DVRs). The cameras can be decentralized, they generally do not have a human monitoring them live, and they do not record continuously – i.e., they can be set off by motion detection or other triggers. In further advances, since 2017, some Internet Protocol (IP)-based cameras have been equipped with software that can conduct automatic biometric recognition, so that, for example, in some Singaporean shopping malls individual humans can be identified [19]. At present there are an estimated 1 billion surveillance cameras in the world [20]. That equates to an average of one surveillance camera per 8 people. In some cities, CCTV cameras outnumber humans 11 to 1 [21].

B. Audiovisual Evidence Manipulation and Deepfakes

In THE CAPTURE there is a team of forensic experts, aptly named "Correction," who are able to manufacture digital evidence when it is unavailable. They can also take existing digital evidence and manipulate it so that there appears to be no reasonable doubt about how to interpret a given event. The process of "correction" ensures that the courts and the jury can find a defendant "guilty." In the series, this evidence tampering is usually conducted using a variety of multimedia techniques that can allegedly go undetected by other stakeholders: everything from deepfakes created by generative AI (GenAI) [22], morphing techniques that bring together two different identities to ensure a biometric match of either individual, advanced creative graphical methods embedding gait in another person's skeletal structure, and the sophisticated editing of images using next generation creative techniques designed for policing and intelligence- tools unavailable to the general public.

Of course, the "Correction" team operates in a morally gray area, insisting that they only "correct" footage through the use of software when they know the person is guilty of committing the crime [23], but where they are not able to

obtain evidence to convict through direct eyewitness accounts or through a warrant process for wiretaps and home searches, or from other kinds of surveillance techniques. The strength of this series is in the possibilities it points to, and already "[r]esearchers are concerned that the same technologies could be used to construct alibis or fabricate criminal evidence in scalable and inexpensive ways … Generative AI poses potential threats, especially in the realm of generating fake evidence or alibis" [24].

C. Cameras Everywhere and Fields of View

Viewers of Episode 1 of THE CAPTURE will be left wondering whether convicting Emery of Roberts's murder is a just outcome. Regardless of whether the viewer concludes it is the *right thing to do* in this fictitious scenario, might the events portrayed in the episode point to a future that requires all of us to wear pin-hole cameras with 360-degree views to ensure an alibi and our own counterevidence? This would be a so-named "live" Jiminy Cricket [25] that broadcasts securely to the Web to prove our guilt or uphold our innocence.

What might such a future mean for self-correction, or the prevention of crime? Might people reform if they know they are creating evidence through lifelogging applications? Here we are reminded of the promises of the Metaverse that will be conducting full body mapping and collecting other private and personal details that may well be used to support the conviction of crime [26]. For many, the Internet, and later social media, were the first forms of near real-time data collection on humans. It is also well-known that Facebook and Instagram, among other leading social media platforms, have become the cheapest investigative tools in the crime solving business, now considered significant forms of OS-INT. But of course, surveilled data is not all authentic, and corroborating different fields of view (FoV) will be necessary in the future when, for example, there is more than one source of CCTV at the scene of a crime. Different FoVs will possibly even create conflicting evidence [27].

While the idea seems new in THE CAPTURE, it is really a reversal of the plot of THE FINAL CUT (2004) starring the late Robin Williams [28]. Williams plays "The Cutter" who removes evidence of crimes that have occurred from historical video recordings, retained in the memories of deceased persons [29]. This is done so that their reputations remain intact posthumously. In THE FINAL CUT, crimes are removed from recorded video evidence stored on a memory implant that have been captured for feature-length memorials viewed at funerals, demonstrating that the Correction Team in THE CAPTURE could work to correct or misinform. As with any technology, "dual use" [30] can point to a technique that can be used to "correct" towards a necessary conviction, or one that can be used to evade conviction. If forensics experts can model things that have not occurred in the natural world, then they surely can act to remove evidence captured in the digital world with even greater ease.

One need only consider what is currently occurring on the Internet to extrapolate what might happen if such software got into the hands of the masses. Might the Internet be flooded by sanitized footage where a wrongdoing once had occurred but was augmented? Might historical files emerge with superfluous doctored scenes that have not occurred, causing confusion about actual real-world events? What is truth? How can we be certain of what we see [31]?

IV. Audiovisual as Evidence

We have precedent in the collection of DNA and its use as evidence in a court of law. Since the inception of gathering DNA, and the availability of techniques to analyze it, admissibility of DNA evidence has been linked to the way evidence is stored and collected. Procedures demonstrating a "chain of custody" for DNA evidence have now been created [32]. In the same way, a digital chain of custody will need to be presented for forensic digital evidence [33], perhaps using a blockchain process. Development of these procedures has already begun. For example, ISO/IEC 27037:2012 provides "guidelines for specific activities in the handling of digital evidence, which are identification, collection, acquisition and preservation of potential digital evidence that can be of evidential value" [34].

In some ways, the movie MINORITY REPORT (2002) pointed to the bottom line, which is that someone needs to be found guilty through legitimate means, not "because some group of people believe it" [35] and are able to manufacture evidence in support of their belief. In MINORITY REPORT the Precogs "don't see what you intend to do, only what you will do." But even then, the existence of a 'minority report' produced by a single Precog who foresees a divergent future, from a different FoV, points to the fallibility of audiovisual evidence in criminal matters [36, p. 108]. MINORITY REPORT ultimately emphasizes that natural justice must be served, but where the legal system fails to adjudicate appropriately, as in THE CAPTURE, a person's conscience might ultimately get the better of them.

In Episode 6 of Season One of THE CAPTURE, Emery is blackmailed by U.K. and U.S. intelligence when they adopt 3D modelling to incriminate him and make him look guilty. He succumbs to the blackmail knowing full well he did not murder Roberts but that he did in fact kill the Afghan soldier. In this episode, knowing he got away with the Afghan murder on a "technicality," he ultimately takes the rap for the Roberts murder because of a guilty conscience. Viewers are left questioning whether "Correction" is indeed a necessity, and whether the outcome can always be this neat.

If we believe what is depicted in THE CAPTURE is an example of technology's looming impact on society, what might this mean for the proliferation of cameras in the context of Smart Cities and how the data being gathered there might be used? Smart Cities not only have image sensors affixed on infrastructure, but have multiple additional sensors that gather audio, among other data types. Tampering can surely be detected post collection, unless it is edited "on the fly" in real-time as it is captured and stored to a personal device or to the Cloud [37].

A. Direct Evidence and Circumstantial Evidence

In light of recent innovations, we are confronted with what this might all mean with respect to the use of audiovisual

evidence in a court of law. Courtroom evidence can take on two forms: (1) eyewitness accounts that provide testimony about something that someone has personally seen or heard, described as "direct evidence;" or (2) "circumstantial evidence," that is, evidence of circumstances that can be relied upon not as being fact directly, but instead pointing to a fact [38]. The question is what category of evidence does CCTV surveillance footage fall under: is it direct evidence or circumstantial evidence? While both direct and circumstantial evidence are generally admissible against a defendant in a legal dispute, so long as that evidence is relevant, does not violate any other rules of evidence and is accompanied by appropriate jury direction [39], [40], direct evidence is considered less problematic because it does not require the finder of fact – whether judge or jury – to fill the gaps in a case by drawing "reasonable" inferences. For this reason, direct evidence – particular in audiovisual form – is highly persuasive. If CCTV footage were to capture the scene of the crime it would be considered, as a matter of law, direct evidence [41]. However, does a camera, like a human, really have the ability to claim the status of offering a direct eyewitness account, or should the audiovisual data from a camera be considered a form of circumstantial evidence that contributes to inference rather than fact [42]?

B. The Problems With CCTV Footage as "Evidence"

Most CCTV footage is captured by infrastructure owned and operated by city, public, or private organizations. Police forces generally do not fund or operate such assets but rely on third parties to provide access on a need's basis, "on-demand", usually accessed shortly after an event has taken place [43]. For these reasons audiovisual surveillance is often seen as infallible. But it is subject to shortcomings. First and foremost, most CCTV only captures video. Second, some CCTV only captures still shots in the form of images "frame by frame" and only in black and white. Third, even if video is continuous, the feed only has a given FoV and can be subject to other defects such as poor lighting, obstruction, or other shortcomings. Fourth, audiovisual footage stemming from a mobile device such as a smartphone, in-car camera, or drone, may not offer a uniform and consistent perspective.

One of the problems associated with audiovisual footage are gaps in the recording, either due to line-of-sight issues, or because the activity in the event partially exits the FoV limitations of a camera. From the perspective of the prosecution, any details that miss key acts of provocation by an offender may negatively impact a clear judgment on a given case. Yet the aim of using CCTV is to reduce or altogether remove reasonable doubt [44].

CCTV footage usually, but not always, provides over-sight and does not capture images at ground level like body-worn cameras. Recording devices are hoisted onto a lamppost, building wall, or fixed structure, providing a wider FoV. At times, dependent on the context, CCTV footage may omit defensive movements by a plaintiff, or victim, or aggressive outcomes or actions by a defendant.

C. The Process of Doctoring Evidence and Spoliation

So, while seen as infallible, either damning or exonerating suspects, video surveillance evidence is not foolproof. This is despite the fact that manufacturers of more advanced CCTV deployments today claim to be able to conduct real-time facial profiling through firmware in the CCTV camera itself. When images from CCTV are taken at nighttime in dimly lit areas or wet conditions, they may be blurry and require forensic experts to carry out the identification of suspects, using advanced facial and body mapping techniques. But the intervention of technology experts can increase the potential for error because experts are interpreting an AI-interpolated overlay, rather than raw footage. When a prosecution team cannot prove that the defendant is guilty "beyond a reasonable doubt," then the raw footage gathered of the case in question may require "correction."

Until recently, the doctoring or fabrication of video footage required significant time, money and know-how. However, with the rapid improvements in the accuracy, speed and volume of deepfake technology and the development of text-to-video GenAI applications, like SORA, doctored videos can now be made by the masses with a few clicks of a smartphone app or website [45]. More than just altering or editing existing footage, deepfake technologies utilize AI and machine learning programs capable of generating increasingly realistic audiovisual material from scratch [46], and augmenting video footage in real-time [47]. These socio-technical advancements cast new light on the "fictitious" acts of correction depicted in THE CAPTURE, raising questions as to the current legalities of this conduct.

V. Legal Implications of Audiovisual Innovation

Having considered the socio-technical plausibility of the central premise of THE CAPTURE, which draws on existing biometric capabilities, the final section of this paper will explore its legal implications. There are a number of challenges for law associated with the use of biometrics to facilitate AI-corrected, augmented or generated video footage. Among these challenges are data privacy, copyright/ intellectual property, protection from discrimination and image-based sexual abuse. We ponder the consequences for some of the hallmarks of natural justice throughout the West: the presumption of innocence, the fair trial, and lawyers' ethical duties of competence and to the administration of justice.

A. The Presumption of Innocence

The presumption of innocence is one of the few doctrines of criminal law enacted across diverse legal systems, whether hailing from a civil law or common law tradition. Enshrined in article 14(2) of the *International Covenant on Civil and Political Rights* (*ICCPR*), the presumption of innocence falls squarely within the ambit of the rule of law, which aims to guard individuals against the arbitrary use of governmental power [48]. Conventionally understood, the right to be presumed innocent until proven guilty is shaped by two constitutive elements: the first, imposing on the prosecution the burden of proving that a defendant is guilty of a criminal

offence (the burden of proof) and, the second, guaranteeing that no guilt can be assigned to the defendant until the criminal offence has been proved beyond reasonable doubt (the standard of proof) [49]. The underlying premise of this human and legal right is Blackstone's ratio, which posits that it is "better that ten guilty persons escape than that one innocent suffer" [50]. Interestingly, as reflected by the presence of the Correction team in THE CAPTURE, there has been a gradual weakening in the public's attitude toward the principle across a vast array of legal jurisdictions [51, p. 23]. This is despite the continued status of the presumption as axiomatic in the criminal justice system.

The central premise of THE CAPTURE subverts the role of the presumption of innocence, by legitimizing the removal of reasonable doubt, through technological means, to reduce the risk of a false-negative outcome at trial. Of course, if GenAI or deepfake technology were to be used for a similar, or more sinister purpose, this too would undermine the presumption of innocence and arguably constitute a breach of international and (in most jurisdictions) domestic public law. But what should we make of existing practices of CCTV "correction" by technology experts, or "predictive policing" that relies on data analytics to identify targets for police intervention, or "machine evidence" that, for example, utilizes AI systems to identify victims, perpetrators and locations of crime, or the use of AI systems by the judiciary to assess the flight risk of defendants in bail proceedings [52]? Pursuant to the presumption of innocence, Article 5(1)(d) of the European Union's (EU) recently published Artificial Intelligence Act (AIA) prohibits the use of AI-based predictions to identify and target individuals (dubbed "predictive identification"), in the absence of "a reasonable suspicion of that person being involved in a criminal activity based on objective verifiable facts and without human assessment thereof" [53], recital 42]. However, such a prohibition (once it comes into effect) will only attach to predictive identification practices within the EU and is likely to apply only in the context of automated processing (articles 2 and 3(52), respectively); thus, discounting the impact of "automation biases" that persist notwithstanding the requirement for their being a "human in the loop" [54].

Whilst promising to enhance the efficiency, efficacy and objectivity of criminal procedures, these emerging innovations in the investigation and prosecution of crime are emblematic of a broader shift from a post-crime to pre-crime society and an accompanying culture of control [55]. The concern, from the perspective of the presumption of innocence is the scientific sophistication and opacity of AI systems, which – unlike DNA evidence before it – operate in a manner that is only *partly* known and not always predictable by their human designers [56]. This information asymmetry, considered in light of the persuasive nature of audiovisual evidence and the effects of "automation bias" on decision-makers, means that the use of AI-powered audiovisual evidence and forensic analysis risks undermining the central elements of the presumption of innocence. This is because a defendant presented with sophisticated AI-generated evidence, including AI-interpolated audiovisual depictions of their alleged malfeasance, may in practical terms be subject to a higher – technically insurmountable - "innocence threshold", correspondingly reducing the State's burden and standard of proof in criminal proceedings [57]. As emerging technologies accelerate the shift from a post-crime to pre-crime society, and enable an increase in criminal conviction rates, it is imperative that we engage in public debate and question whether it is indeed better to allow ten innocent persons to suffer in order to prevent one crime.

B. The Fair Trial

Beyond the presumption of innocence, natural justice demands the opportunity for an accused to be heard as embodied by the right to a fair trial in article 14(3) of the *ICCPR*. The ultimate objective of a fair trial is to arrive at the truth. In the context of the adversarial system of justice operating in common law jurisdictions such as the U.S., U.K. and Australia, this truth is said to emerge from a competitive battle between two opposing parties to a dispute adjudicated by an impartial judge or jury. Assuring the fairness of this battle, are rules of evidence that limit the evidence adduced at trial to that which is relevant, authentic, more probative than prejudicial and not improperly obtained [40], rr 101, 102, 401, 403, 901, [58].

On the one hand, advances in AI technologies may enhance the right to be heard by addressing power imbalances within the adversarial pursuit of truth. A great promise of GenAI is its ability to improve access to justice by, for example, developing chatbots that can freely and readily answer legal questions and direct users to better-tailored information and services [59]. The availability of ChatGPT, automated chatbots, and Online Dispute Resolution (ODR) systems for certain types of civil claims [60], has implications for underserved communities or those who cannot afford high quality legal representation [61], pp. 40–47]. On the other hand, the AI technologies at the focus of this paper – capable of generating or manipulating audiovisual "evidence" – present unique challenges to key components of a fair trial, including existing rules for authenticating evidence and, conversely, excluding evidence on the grounds of tampering or falsification.

1) Deepfake Evidence Entering the Court Room: In Part IV, we discussed the complications associated with relying on audiovisual evidence, such as CCTV footage, as direct evidence in court proceedings. Here, we are concerned more explicitly with the admissibility of audiovisual evidence on the grounds of its authenticity. Ascertaining the authenticity of evidence is an issue that courts have grappled with long before the phenomena of deepfakes and GenAI. For example, authenticity is a matter relevant to determining whether a copy of traditional documentary evidence is genuine – that it is what it claims to be – and that its condition remains substantially unchanged. Nevertheless, authenticity is particularly significant when determining the admissibility of *electronic* evidence - implicated, for example, when relying on OSINT hosted on social media platforms [62] – due to the fundamentally mutable nature of digital records.

Now consider the impact of AI-manipulated digital evidence, such as deepfakes, entering the court room. When, in episode 1 of THE CAPTURE, Emery is shown CCTV footage

"capturing" him assaulting and kidnapping Roberts near a bus stop, he lashes out in violent protest claiming the footage to be fake. In his subjective version of "truth", he had kissed Roberts before she boarded a bus to go home – now he is in custody accused of her murder. How can seemingly incontrovertible video proof depicting, for example, the accused engaging in an alleged offence or, conversely, at an alternative location at the time of an alleged offence, be challenged in court by lawyers? How can a court handle the proper authentication of such evidence? And, if deepfake video proof is permitted to be seen by a jury, can any *reasonable* doubt as to the accused's guilt remain?

The rapid improvement in deepfake technologies will soon present litigants, lawyers, and judges with dilemmas of this very nature. At least two well-known cases have already been the subject of deepfake evidence [63]. The first case, heard in the U.K., involved a woman who used an audio deepfake against her former spouse in a child custody trial. While the defendant successfully used metadata analysis to prove that the audio file had been falsified, lawyer for the defendant, Byron James, noted that "courts take evidence such as audio recordings, visual footage and written documents at face value" [64]. The second case, heard in the U.S., involved the infamous Spone harassment trial of the dubbed "deepfake cheerleader mom" who was alleged to have created a deepfake video of a peer in her daughter's cheerleading squad vaping and altered the social media accounts of other girls on her daughter's cheerleading squad to place them in compromised contexts [65]. Spone denied creating deepfakes and, in May 2021, the prosecutor's office announced that it was no longer pursuing the deepfake video as the basis for charges as police had been unable to confirm that the video evidence was falsified. Digital forensic experts had determined that the footage appeared to be authentic, not a deepfake, but noted that poor video quality and the lack of other evidence in the case made it impossible to draw a clear conclusion [66]. Although the deepfake evidence was not used against the victims in either of these highly publicized examples, they have been described as "cautionary tales" of the power of deepfakes as a source of false evidence capable of infecting the determination of truth in court proceedings [67]. As deepfake technologies improve, become more readily accessible and outpace the technology used to detect them [68], novel questions arise regarding the efficacy of authenticity requirements within existing laws of evidence.

2) Rules for Authenticating Evidence: Doctrinally, the precise requirements of authentication differ from jurisdiction to jurisdiction. In the U.S., for example, authenticity is an explicit ground for admissibility. Rule 901(a) of the *Federal Rules of Evidence* (FRE) requires the proponent to prove the identity and authenticity of evidence by producing "evidence sufficient to support a finding that the item is what the proponent claims it to be". This can be achieved in a variety of ways including presenting testimony from a witness with personal or expert knowledge, by relying on the distinctive characteristics of the evidence such as the "appearance, contents, substance, internal patterns, or other distinctive characteristics", providing evidence that a process or system "produces an accurate result" [40, rr 901(b)(1), (3), (4), (9)], or providing a certified copy of records "generated by an electronic process or system" [40, rr 902(13)-(14)]. Pursuant to these rules, the authentication of digital videos or images may require human testimony and/or supporting details regarding chain of custody, inclusive of metadata [69]. However, the threshold for establishing a prima facie case of authenticity has been described as "slight" [70, p. 1404], with the proponent of evidence (including electronically stored digital evidence) merely needing to provide "a foundation from which a jury could reasonably find that the evidence is what the proponent says it is" [71, p. 38]. Following this, it is for the jury to make the ultimate determination regarding the authenticity of and, separately, the weight afforded to the evidence [72]. This "non-rigorous" approach persists even though trial judges are increasingly cognizant of the complexity of ascertaining the truth of evidence "in the age of fake social-media accounts, hacked accounts, and so-called deep fakes" [73, p. 565].

In Australia, the *Uniform Evidence Law* (UEL) similarly requires the proponent of evidence to establish the identity and authenticity of evidence, with ss 166-167 permitting authentication via the testimonial evidence of someone with personal knowledge about the "document or thing" and ss 170-172 permitting authentication by way of a statement or affidavit by "a person who, at the relevant time or afterwards, had a position of responsibility in relation to making or keeping the document or thing" [58]. Beyond the reliance on human testimony for the authentication of evidence, ss 146-147 of the UEL provides a rebuttable presumption that evidence produced by a device or process (such as a video recording device) "produced that outcome". This presumption weighs in favor of electronic evidence generated by technological means being considered genuine and reliable without having to prove the accuracy or proper functionality of that technological device. Judicial authority in Australia is unsettled as to whether challenges to authenticity under the UEL are matters of law to be determined by a judge or matters of fact to be determined by jury. Unlike the U.S. FRE, the Australian UEL neither expressly designates nor omits authentication as a condition-precedent to admissibility [75]. Although no source can be found in the Act, judicial commentary suggests that the absence of evidence as to the provenance of a "document or thing" *may* provide a basis for its *discretionary* exclusion by a court under s 135. Nevertheless, as Caruso, Legg and Phoustanis have emphasized in the context of data deriving from eObjects, the outcome of this debate is only relevant to the question of *who* determines an authentication issue; it does not address the pertinent issue of *how* challenges to authentication can be resolved given the complexities associated with AI generated or interpolated evidence [76].

3) Detecting Deepfake Evidence: Across both jurisdictions there are no rules of evidence, or evidentiary procedures, that explicitly govern the presentation of GenAI or deepfake evidence in court. In fact, the legal standards for authenticating evidence have remained largely unchanged throughout 21st century advancements in digital technologies - from capturing and storing media on smart phones to readily posting and sharing content on social media platforms or dabbling in

"Photoshopping" - likely because the creation of convincing fake evidence was prohibitively difficult, and courts could rely on expert witnesses where authenticity was in question [77]. Indeed, the detection of deepfake evidence is a matter ignored by the plot of THE CAPTURE. Afterall, couldn't the tampering of CCTV footage in the Emery case be detected by forensic experts using time and date stamps among other metadata captured at the time of recording? If police were the party doing the spoliation - by confiscation, destruction, or manipulation of video or photographic evidence - could they not be prosecuted for evidence tampering?

The problems with applying existing rules of authentication, grounded in a human-centric paradigm, to AI-generated or augmented evidence are manifold. First, as deepfake technology has advanced it has become increasingly difficult to detect with the naked eye [78], with even experts struggling to detect the veracity of potential deepfakes [79]. Furthermore, it has long been known that watching fake audiovisual depictions of an event is likely to corrupt witness memory, in part, because humans value visual perception above other indicators of truth [80]. In combination, these factors undermine the ability to rely on human testimony to authenticate visual evidence in the age of sophisticated deepfake technology. These same factors will impact the decision-making of jurors in jurisdictions where the question of authenticity is left to the tribunal of fact. Second, and somewhat conversely, as public awareness of the nature, quality and prevalence of deepfakes increases, so too will their skepticism regarding the authenticity of "real" audiovisual evidence – illustrated by the allegations made in the Spone harassment trial. This concern has been dubbed the "liar's dividend" by Chesney and Citron who point to the increased ease with which liars can deny the truth in the age of deepfake technology [81]. This skepticism may be preyed upon by lawyers wishing to persuade a jury to accord little weight to "real" video footage that is highly damaging to their client's case. More generally, it may lead to diminished judicial (both judge and jury) confidence in audiovisual evidence contributing to false-negative outcomes at trial. The "liars' dividend" has implications for deciding whether (and how) to amend existing rules of evidence for the purposes of raising the authentication threshold.

The third challenge facing existing rules of evidence pertains to the difficulties in relying on the forensic analysis of technology experts or technological methods, such as "verified media capture technologies" [82], for detecting AI- generated or augmented evidence. One hurdle in this challenge concerns the sheer cost and time involved in engaging technology experts to assess the authenticity of audiovisual evidence. If deepfake detection requirements were baked into the rules of evidence, as a response to the prevalence of AI manipulation and plummeting confidence in believing what we see, the commensurate rise in litigation costs would cause renewed concerns for access to justice. An issue of more immediacy, however, is the reality that deepfake technologies created through GANs or DMs persistently outpace the techniques and programs used to detect them. As John Ruff aptly notes "[a]ny program, purporting to identify a new way to detect fake videos can simply be incorporated into the GAN cycle by the creators of deepfakes, rendering the detection model obsolete" [83]. Paradoxically the net result of research aimed at *detecting* deepfakes, is *more* powerful deepfake technology; particularly in the context of DMs capable of achieving imperceptible start-of-the-art performance [84].

C. Lawyers, Ethics and Skills

Notwithstanding the limitations of criminal sanctions as complete deterrents, the challenges in implementing effective detection and authentication strategies for AI-generated or augmented evidence underscore the need for substantive laws to proscribe the use of harmful deepfake content such that its entry in the court room would be consequently curbed. While in June 2024 the Australian Government introduced a Bill to create new criminal offences to ban the sharing of deepfake pornography [85], and in January 2024 a Bill was introduced in the U.S. Senate to regulate a suite of deepfake activities [86], the ultimate safeguard against the use of AI-manipulated evidence in the *courtroom* is the legal profession. As such, in addition to challenging the procedural requirements of the presumption of innocence and laws of evidence, advances in GenAI and deepfake technologies place a greater responsibility on lawyers, in the first instance, to question and challenge the authenticity of audiovisual evidence, and judges, in the second, to understand the nature and complexities of these emerging socio-technical practices. In this section of the paper, we briefly explore the contradictory implications of GenAI and deepfake audiovisual evidence on two fundamental ethical duties of legal practitioners - (1) the duty of competence and (2) the duty to the administration of justice – and contemplate what this means for the future of legal education and continued skill development.

1) The Ethical Duties of Lawyers to the Client: Competence and Diligence: An obvious quandary that arises in the context of AI-manipulated or deepfake evidence is *how* this evidence enters the court room. Are lawyers, representing litigants in legal disputes, legally and ethically permitted to use such evidence? Would the District Attorney prosecuting Emery's case in THE CAPTURE be permitted to adduce falsified incriminating evidence with impunity?

As professionals with a monopoly on legal services, lawyers are bound by ethical duties enforced by legislation; breach of which could constitute professional misconduct resulting, in the most severe circumstances, in disbarment from the profession [87, r 8.4], [88, r 2.3]. While the exact content and scope of the framework governing legal ethical responsibility might differ from jurisdiction to jurisdiction, the fundamental duties of lawyers can be broadly characterized by obligations to the client, on the one hand, and obligations to the court, on the other, with the duty of honesty – to courts, clients and other legal practitioners – functioning as the golden thread that weaves between [89]. Advances in GenAI and deepfake technologies implicate the ethical responsibilities of lawyers in somewhat divergent ways, with some obligations limiting or forbidding its use and other obligations increasingly necessitating it.

With respect to lawyers' obligations to the client, both the American Bar Association's *Model Rules of Professional Conduct* (MRPC) adopted by a majority of U.S. States and the *Australian Solicitor's Conduct Rules* (ASCR) governing a majority of Australia's practicing lawyers, require lawyers to act in the best interests of their clients. This encompasses the duty to zealously serve clients with both competence and diligence [87, rr 1.3-1.4, 1.6], [88, r 4.1.1-4.1.3, 8-9]. Underlying lawyers' ethical duties to their clients, throughout common law jurisdictions, are fiduciary obligations - of loyalty, no-conflict and no-profit - that attach to the lawyer-client relationship, which is characterized in law as a "relationship of trust and confidence" [90]. Accordingly, a lawyer's duty to their client requires them to pursue their client's interests before their own.

The relationship between emerging technologies and lawyers' ethical duties to their clients has been explicitly recognized in the U.S. since 2012. This is a result of the enactment by the American Bar Association of comment 8 to MRPC rule 1.1, which explains that "[t]o maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology…". To date, 40 out of 50 U.S. states have adopted the duty of technology competence [91], which has been used as grounds for disciplinary action against lawyers who, for example, have failed to meet electronic pleadings requirements [92] or failed to properly provide electronically stored information (ESI) during electronic discovery [93]. With respect to AI more specifically, in April 2024 the New York State Bar Association's (NYSBA) Task Force on Artificial Intelligence noted that the duties of competence and diligence ought to require lawyers to actively consider the benefits for both quality and efficiency of using AI and GenAI when providing legal services to a specific client [61].

In jurisdictions outside of the U.S. where there is no legally recognized duty of technology competence, such as Australia, existing ethical duties to the client may nevertheless require lawyers to update their knowledge and use of emerging technologies. As legal practice has become infused with AI technologies (dubbed "LawTech") – from the use of Technology Assisted Review (TAR) software for document discovery, to the deployment of data analytics to predict the outcome of potential litigation, and the use of AI to automate document drafting [94] – it has become incumbent upon lawyers to adopt AI tools that improve the quality and/or efficiency of their work. This is because, for many systematic tasks, such technology "can produce more accurate results, for less costs, and in a much quicker timeframe" [95, p. 466], therefore, enhancing a lawyer's compliance with their obligations of competence and diligence. Indeed, the uptake of LawTech across the public and private sector is driven not by the innovation of lawyers, but by the demand of clients who expect costs to be controlled [96, p. 255]. It would be unethical, therefore, for lawyers to charge excessive fees that result from undertaking manually those legal tasks that could be readily completed using reliable cost-saving AI [97].

2) The Ethical Duties of Lawyers to the Court: Facilitating the Administration of Justice: The pertinent question, however, is how do lawyers' ethical duties intersect with the use of AI-manipulated or deepfake audiovisual evidence? Do lawyers' duties of competence, diligence and loyalty to the client permit the use of outcome-enhancing deepfake evidence upon a client's instruction? While lawyers are required to act on clients' instructions, they cannot act on unlawful instructions [87, r 8.1], [88, r 1.2(d)]. And while it may not be *expressly* unlawful for a client to suggest the use of AI-manipulated or deepfake evidence, counterbalancing a lawyer's duties to the client is their duty to the court and, correspondingly, the rule of law and the administration of justice. As put by Littrich and Murray, a lawyer is not simply a "hired gun" [98, p. 243].

In Australia, unlike the U.S., the duty to the court is expressly designated as the *paramount* duty of lawyers, which "prevails to the extent of inconsistency with any other duty" [88, r 3.1]. Regardless of its primacy, however, the duty to the court prohibits lawyers in either jurisdiction from knowingly or recklessly deceiving or misleading the court by offering deepfake evidence [87, rr 3(a)(1), 3(a)(3)]; [88, rr 4.1.2, 4.1.4, 5.1, 19.1]. Comment 6 to MRPC rule 3.3(a)(3) clarifies that a lawyer must refuse to offer false evidence, such as a deepfake video, even if they are insisted to do so by a client. By extension, a lawyer will have breached their ethical duties to the court if they at first unknowingly offer deepfake evidence, but subsequently learn of its falsehood and fail to inform the court prior to judgment [87, r 3(a)(3)]; [99]. Conversely, in the absence any indication of "obvious falsehood", a lawyer will *not* knowingly or recklessly mislead or deceive a court by offering evidence that they reasonably believe to be false [87, r 3(a)(3), ct 8].

Moreover, lawyers' ethical duties to the court are also implicated when seeking to challenge the authenticity of the opposing party's evidence. Querying and excluding AI-manipulated or deepfake evidence aligns with a lawyer's duty to the court in so far as it preserves the integrity of the judicial process and facilitates the administration of justice. However, such challenges may contribute to a jury's skepticism regarding the authenticity of *real* evidence, not only risking a miscarriage of justice, but undermining public confidence in the truth-finding function of the court. For these reasons, Pfefferkorn notes that lawyers "must tread carefully when weighing whether to accuse the other side's evidence of deepfakery" [100, p. 274].

3) Legal Skills and Education: The intersection between the rapid advancement in deepfake technologies and lawyers' ethical professional responsibilities presents a number of unique challenges for lawyers and the future of legal education. On the one hand, it is decisively unethical under existing rules of conduct to knowingly offer false evidence to the court. Thus, unless legislation authorized the correction of murky evidence by a "special" government agency, it would be currently unlawful for a district attorney or state prosecutor to offer in court the doctored footage depicted in THE CAPTURE. On the other hand, a suspicion or reasonable belief that audiovisual evidence has been the subject of AI-manipulation or deepfakery does not prevent a lawyer from offering it in support of their client's case and we may see the rise in lawyer's challenging the authenticity of "real"

audiovisual artefacts presented in court, given the ease with which a factual basis for such a challenge could be provided in the deepfake age. But what ought to be the consequence for a lawyer who unknowingly relies on deepfake evidence? Unlike other innovations in forensic evidence - from DNA to digital recordings or social media posts – deepfakes produced through GANs or DMs, by design, trick the viewer. How does the modern lawyer satisfy their duties of competence, diligence and candor in the era of AI-powered technologies that raise significant existential questions about reality and, indeed, the existence of a distinction between fact and fiction?

The consequences for lawyers lacking sufficient understanding of emerging GenAI technologies are evident in the use of ChatGPT – an example of a Gen-AI, large language model (LLM) chatbot. A hallmark of the skill required of lawyers is the ability to identify and understand legal principles and engage in legal reasoning to apply those principles to a vast array of problems. The advent of GenAI, however, tests these traditional legal skills. All around the world, lawyers (and self-represented litigants) adopting GPT-style engines, without understanding the pitfalls, have been caught out. In the infamous 2023 U.S. case, *Mata v Avianca, Inc.,* the plaintiff's lawyers submitted a brief containing fake judicial opinions, case citations and judicial quotes to the New York District Court [101]. The lawyers insisted on the veracity of these precedents even after judicial orders called the fake judgements into question. ChatGPT's hallucinations, in this instance, were the undoing of the lawyers who were sanctioned and fined for acting in bath faith [102]. In its 2024 report, the NYSBA's Task Force on Artificial Intelligence commented that this case (and a host of others like it) signal that "attorneys cannot rely on technology without verification" [61, p. 29]. Indeed, the combination of a lawyer's duty of competence and duty to the administration of justice, may now require lawyers to have a basic understanding of how the AI programs they use – or are likely to encounter – operate and to not unquestioningly accept their results as true. Of course, this becomes exceedingly difficult as AI advances and is well beyond the bounds of traditional legal education.

This is now a call to the legal, information services, and forensic companies to acknowledge the problems that GenAI and deepfakes pose for our wider community [22], and act to address these through commensurate counter services. This may include rethinking: (1) the nature and scope of legal education; (2) rules of evidence to have the development of a chain of custody for textual, audio and audiovisual evidence; and (3) to information companies and media publication services to ensure that their databases are authenticated and false court hearings do not enter these important information sources, that are increasingly being used by bots and GPTs to extract details in a fraction of the time. In all of this, ethics and accountability play important roles. While no one is questioning the use of emerging technologies for legitimate purposes, there needs to be a human-in-the-loop – with appropriate resources and skills - to counter check the data being generated by GenAIs. Legg and McNamara stress that if this trend continues in the legal profession without correction, GenAI has the potential to "undermine the public's trust in the legal system" [102].

VI. Conclusion

In the end, THE CAPTURE series calls us to think about the impacts of modern technology on society. The falsification of evidence is interfering with justice. While one cannot alter evidence through intentional or reckless fabrication with impunity, GenAI and deepfake technologies are facilitating the creation of highly sophisticated artefacts, such as audiovisual footage, that are designed to trick the most skeptical and judicious of individuals. After examining the potential dilemmas that deepfakes, as a form of disinformation, create for society at large, this paper demonstrated the implications of advances in GenAI for central tenets of natural justice, including (but not limited to): (a) the presumption of innocence, (b) the fair trial, and (c) the ethical duties of lawyers. While eradicating deepfakes altogether is impossible, despite the bans that some jurisdictions have introduced, we call for further research into the notion of a digital chain of custody framework to be developed and observed as a new standard for dealing with real versus fake data. In this way we can still enjoy the benefits of GenAI, GANs and DMs without the commensurate known negative intended consequences of this emerging technology.

Acknowledgment

We would like to thank Executive Editor Sonia Epstein of the Museum of the Moving Image for her edits of the original paper that inspired this new contribution, titled: "Beyond A Reasonable Doubt? Review of THE CAPTURE" published on February 13, 2023 https://scienceandfilm.org/articles/3522/beyond-a-reasonable-doubt-review-of-the-capture.

References

[1] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119, Art 32*, Eur. Union, Brussels, Belgium, 2016.

[2] A. Sharma, "99% adults in India have Aadhaar number and 'use it at least once a month', says UIDAI." Firstpost. Jul. 23, 2022. [Online]. Available: https://www.firstpost.com/india/99-9-adults-in-india-have-aadhaar-number-and-use-it-at-least-once-a-month-says-uidai-10948761.html

[3] S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch, "Face morphing attack generation and detection: A comprehensive survey," *IEEE Trans. Technol. Soc.*, vol. 2, no. 3, pp. 128–145, Sep. 2021. doi: 10.1109/TTS.2021.3066254.

[4] K. Michael, "Leading with evidence: Operationalizing public interest technology in practice," presented at the AAAS Annu. Meeting Empower With Evidence, Feb. 2022. [Online Video]. Available: https://youtu.be/rAhTEc4X2vQ

[5] A. McDowell, "Storytelling shapes the future," *J. Futures Stud.*, vol. 23, no. 3, pp. 105–112, 2019.

[6] L. Perusco and K. Michael, "Control, trust, privacy, and security: Evaluating location-based services," *IEEE Technol. Soc. Mag.*, vol. 26, no. 1, pp. 4–16, Mar. 2007.

[7] K. Michael, J. R. Schoenherr, and K. M. Vogel, "Failures in the loop: Human leadership in AI-based decision-making," *IEEE Trans. Technol. Soc.*, vol. 5, no. 1, pp. 2–13, Mar. 2024, doi: 10.1109/TTS.2024.3378587.

[8] K. Michael, M. G. Michael, R. Abbas, "The importance of scenarios in the prediction of the social implications of emerging technologies and services," *J. Cases Inf. Technol.*, vol. 13, no. 2, pp. 1–7, 2011.

[9] Q. J. Ullrich, "Is this video real? The principal mischief of deepfakes and how the Lanham Act can address it," *Columbia J. Law Social Probl.*, vol. 55, no. 1, pp. 1–56, 2024.

[10] B. Chanan, "The Capture." IMDB. 2019. [Online]. Available: https://www.imdb.com/title/tt8201186/

[11] K. Vellani, "Deepfakes and legal implications: Seeing is not believing." Clifford Chance. Dec. 21, 2020. [Online]. Available: https://www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2020/12/deepfakes-and-legal-implications--seeing-is-not-believing.html

[12] R. Bleiker, "Authenticity and deception in an age of visual wearables," *IEEE Technol. Soc. Mag.*, vol. 33, no. 2, pp. 26–27, Jun. 2014.

[13] G. Voss, "Fraud groups use deepfakes to enhance imitation scams in Peru," Insight Crime, Jul. 21, 2023. [Online]. Available: https://insightcrime.org/news/fraud-groups-use-deepfakes-to-enhance-imitation-scams-in-peru/

[14] H. Chen and K. Magramo, "Finance worker pays out $25 million after video call with deepfake 'chief financial officer.'" CNN. Feb. 4, 2024. [Online]. Available: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

[15] Staff, "'Deepfake' scam in China fans worries over AI-driven fraud," Reuters. May 22, 2023. [Online]. Available: https://www.reuters.com/technology/deepfake-scam-china-fans-worries-over-ai-driven-fraud-2023-05-22/

[16] N. Mohan and E. Chang, Bloomberg Television. *YouTube Says Using Videos to Train OpenAI's Sora Breaks Rules*. (2024). [Online Video]. Available: https://youtu.be/FBZ__BeChRg?si=qBmupoQkH6Oe7nk0

[17] The Wall Street Journal. *OpenAI's Sora Made Me Crazy AI Videos—Then the CTO Answered (Most of) my Questions*. (Mar. 13, 2024). [Online Video]. Available: https://www.youtube.com/watch?v=mAUpxN-EIgU

[18] L. H. Martinson, "Batman: The movie." IMDB. 1966. [Online]. Available: https://www.imdb.com/title/tt0060153/

[19] iOmniscient. *Suntec Singapore Implements iOmniscient's Video Analytics for Better Security and Service*. (2016). [Online Video]. Available: https://www.youtube.com/watch?v=eowNe6FDFm0

[20] Staff, "Who's watching: The cities with the most CCTV cameras," Geographical, Mar. 7, 2023. [Online]. Available: https://geographical.co.uk/science-environment/whos-watching-the-cities-with-the-most-cctv-cameras

[21] B. Sheldon, "Camera surveillance within the U.K.: Enhancing public safety or a social threat?" in *Developments in Counter-Terrorist Measures and Uses of Technology*. London, U.K.: Routledge, 2013, pp. 87–97.

[22] S. Karnouskos, "Artificial intelligence in digital media: The era of deepfakes," *IEEE Trans. Technol. Soc.*, vol. 1, no. 3, pp. 138–147, Sep. 2020.

[23] E. B. Griffiths, "The Capture explained: What really happened?" RadioTimes, Oct. 1, 2019. [Online]. Available: https://www.radiotimes.com/tv/drama/the-capture-episode-5-reveals-what-really-happened/

[24] E. Ferrara, "GenAI against humanity: Nefarious applications of generative artificial intelligence and large language model," *J. Comput. Soc. Sci.*, to be published. [Online]. Available: https://doi.org/10.1007/s42001-024-00250-1

[25] "Jiminy Cricket," Wikipedia, May 20, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Jiminy_Cricket

[26] K. Michael, R. Abbas, and S. Papagiannidis, "The social implications of XR: Promises, perils, and potential," *IEEE Technol. Soc. Mag.*, vol. 43, no. 1, pp. 91–108, Mar. 2024, doi: 10.1109/MTS.2024.3370023.

[27] K. Hamann and R. R. Brown, "Secure in our convictions: Using new evidence to strengthen prosecution," *Criminal Just.*, vol. 32, no. 4, pp. 25–31, 2018.

[28] O. Naim, "The Final Cut." IMDB. 2024. [Online]. Available: https://www.imdb.com/title/tt0364343/

[29] K. Michael, "The Final Cut—Tampering with direct evidence from wearable computers," presented at the 5th Int. Conf. Multimedia Inf. Netw. Security (MINES), Beijing, China, 2013.

[30] M. J. Selgelid, "Governance of dual-use research: An ethical dilemma," *Bull. World Health Org.*, vol. 87, no. 9, pp. 720–723, 2009.

[31] T.L. Wagner and A. Blewer, "'The word real is no longer real': Deepfakes, gender, and the challenges of AI-altered video," *Open Inf. Sci.*, vol. 3, no. 1, pp. 32–46, 2019.

[32] K. Michael, "The legal, social and ethical controversy of the collection and storage of fingerprint profiles and DNA samples in forensic science," in *Proc. IEEE Int. Symp. Technol. Soc.*, Wollongong, NSW, Australia, 2010, pp. 48–60, doi: 10.1109/ISTAS.2010.5514654.

[33] "Standards and best practices for digital forensics," United Nations Office on Drugs and Crime, Mar. 2019. [Online]. Available: https://www.unodc.org/e4j/en/cybercrime/module-4/key-issues/standards-and-best-practices-for-digital-forensics.html

[34] *Information Technology—Security Techniques—Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence*, ISO/IEC Standard 27037:2012, 2012. [Online]. Available: https://www.iso27001security.com/html/27037.html

[35] S. Spielberg, "Minority Report." IMDB. 2002. [Online]. Available: https://www.imdb.com/title/tt0181689/

[36] C. Sharp, "Adversarial justice: Truth, trials and juries" in *Law and Popular Culture in Australia*, B. Richards, M. de Zwart and S. Le Mire, Eds., Chatswood, NSW, Australia: LexisNexis, 2015, pp. 99–113.

[37] J. Kahn, "Business execs are facing a new threat: The person talking on Zoom might be an A.I.-generated 'deep fake,'" Fortune, Sep. 3, 2022. [Online]. Available: https://fortune.com/2022/09/03/live-deepfakes-detect-methods-zoom-fraud/

[38] "4.02 direct and circumstantial evidence defined," New York State Unified Court System, May 20, 2024. [Online]. Available: https://www.nycourts.gov/JUDGES/evidence/4-RELEVANCE/4.02_Direct_and_Circumstantial_Evidence_Defined.pdf

[39] Evidence Act 1995 (Cth), Sep. 2021. [Online]. Available: https://jade.io/j/?a=outline&id=216645

[40] Federal Rules of Evidence, Art IV, Art IX, 2023. [Online]. Available: https://uscode.house.gov/view.xhtml?path=/prelim@title28/title28a/node230&edition=prelim

[41] People v Hardy, 26 NY3d 245, 251 (Court of Appeals of New York), 2015. [Online]. Available: https://case-law.vlex.com/vid/people-v-hardy-no-885952974

[42] U. Nedim, "What are the problems with using CCTV evidence in court?" NSW Courts. Oct. 13, 2014. [Online]. Available: https://nswcourts.com.au/articles/what-are-the-problems-with-using-cctv-evidence-in-court/

[43] "CCTV register," NSW Police, May 20, 2024. [Online]. Available: https://www.police.nsw.gov.au/__data/assets/pdf_file/0011/105113/nsw_cctv_brochure.pdf

[44] F. Dahlstrom, "What does 'beyond a reasonable doubt' mean?" Go to Court. Jun. 5, 2019. [Online]. Available: https://www.gotocourt.com.au/criminal-law/beyond-a-reasonable-doubt/

[45] R. Riyan, Jun. 7, 2024, "The best deepfake apps and websites in 2024," Companion Link Software. [Online]. Available: https://www.companionlink.com/blog/2024/05/the-best-deepfake-apps-and-websites-in-2024/

[46] eSafety Staff, "Deepfake trends and challenges—Position statement," Australian Government eSafety Commissioner, Jan. 23, 2022. [Online]. Available: https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes

[47] W. Jackson, "How South-East Asia's pig butchering scammers are using artificial intelligence technology," ABC News. May 15, 2024. [Online]. Available: https://www.abc.net.au/news/2024-05-16/pig-butchering-scams-artificial-intelligence-ai-face-swapping-/103804830

[48] M. Krygier, "The rule of law: Pasts, presents and two possible futures," *Annu. Rev. Law Social Sci.*, vol. 12, pp. 199–299, Aug. 2016.

[49] Human Rights Committee, General Comment No. 32, Article 14: Right to equality before court and tribunals and to a fair trial, 19th session, United Nations, New York, NY, USA, document CCPR/C/GC/32, Aug. 2007.

[50] W. Blackstone, *Commentaries on Law*, William & Mary, Williamsburg, VA, USA, 1975.

[51] M. Xiong, R. C. Greenleaf, and J. Goldschmidt, "Citizen attitudes toward errors in criminal justice: Implications of the declining acceptance of blackstone's ratio," *Int. J. Law, Crime Just.*, vol. 48, pp. 14–26, 2016.

[52] A. Završnik, "Criminal justice, artificial intelligence systems, and human rights," *ERA Forum*, vol. 20, pp. 567–583, Feb. 2020.

[53] *Proposal for a Regulation (EU) 2024/... of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013,) EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, Eur. Union, Brussels, Belgium, 2018.

[54] U. Agudo, K. G. Liberal, M. Arrese, and H. Matute, "The impact of AI errors in a human-in-the-loop process," *Cogn. Res. Principles Implicat.*, vol. 9, no. 1, pp. 1–16, 2024. [Online]. Available: https://doi.org/10.1186/s41235-023-00529-3

[55] L. Strikwerda, "Predictive policing: The risks associated with risk assessment," *Police J. Theory, Pract. Principles*, vol. 12, no. 3, pp. 422–436, 2020.

[56] A. Sachoulidou, "Algorithmic criminal justice: Is it just a science fiction plot idea?" in *Artificial Intelligence and Normative Challenges* (Law, Governance and Technology Series), vol. 59, A. Kornilakis, G. Nouskalis, V. Pergantis, T. Tzimas, Eds. Cham, Switzerland: Springer, 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-41081-9_8

[57] A. Sachoulidou, "Going beyond the 'common suspects': To be presumed innocent in the era of algorithms, big data and artificial intelligence," *Artif. Intell. Law*, to be published. [Online]. Available: https://doi.org/10.1007/s10506-023-09347-w

[58] *Evidence Act 1995* (Cth) SECT 55-58, 135-8, 1995. [Online]. Available: https://classic.austlii.edu.au/au/legis/cth/consol_act/ea199580/

[59] S. La Roque-Doherty, "Artificial intelligence has made great inroads, but hasn't yet increased access to civil justice." ABA Journal. Apr. 1, 2024. [Online]. Available: https://www.abajournal.com/magazine/article/artificial-intelligence-has-made-great-inroadsbut-not-as-far-as-increasing-access-to-civil-justice#google_vignette

[60] S. Salter, "Online dispute resolution and justice system integration: British Colombia's civil resolution tribunal," *Windsor Yearbook Access to Just. J.*, vol. 34, no. 1 pp. 112–129, 2017.

[61] *NYSBA Task Force on Artificial Intelligence: Report and Recommendations of the New York State Bar Association Task Force on Artificial Intelligence*, New York State Bar Assoc., Albany, NY, USA, Apr. 2024.

[62] People v Beckley, 185 CalApp 4th 509 (California Court of Appeal), 2020. [Online]. Available: https://www.courtlistener.com/opinion/2265148/people-v-beckley/

[63] R. A. Delfino, "Deepfakes on trial: A call to expand the trial judge's gatekeeping role to protect legal proceedings from technological fakery," *Hastings Law J.*, vol. 74, no. 2, p. 293, 2023. [Online]. Available: https://repository.uchastings.edu/hastings_law_journal/vol74/iss2/3

[64] G. Swerling, "Doctored audio evidence used to damn father in custody battle," The Telegraph, Jan. 31, 2020. [Online]. Available: https://www.telegraph.co.uk/news/2020/01/31/deepfake-audio-used-custody-battle-lawyer-reveals-doctored-evidence/

[65] K. Bellware, "Cheer mom used deepfake nudes and threats to harass daughter's teammates police say," The Washington Post, Mar. 13, 2021. [Online]. Available: https://www.washingtonpost.com/nation/2021/03/13/cheer-mom-deepfake-teammates/

[66] D. Harwell, "Remember the 'deepfake cheerleader mom'? Prosecutors now admit they can't prove fake-video claims." The Washington Post. May 14, 2021. [Online]. Available: https://www.washingtonpost.com/technology/2021/05/14/deepfake-cheer-mom-claims-dropped/

[67] M. Reynolds, "Courts and lawyers struggle with growing prevalence of deepfakes." ABA Journal. Jun. 9, 2020. [Online]. Available: https://www.abajournal.com/web/arti-cle/courts-and-lawyers-struggle-with-growing-prevalence-of-deepfakes

[68] N. I. Brown, "Deepfakes and the weaponization of disinformation," *Virginia J. Law Technol.*, vol. 23, no. 1, pp. 23–82, 2020.

[69] Lorraine v. Markel American Ins. Co., 241 F.R.D. 534 (D. Md. 2007). 2007. [Online]. Available: https://casetext.com/case/lorraine-v-markel-american-ins-co

[70] United States v Reilly, 33 F.3d 1396 (United States Court of Appeals). 1994. [Online]. Available: https://casetext.com/case/us-v-reilly-5

[71] United States v Safavian, 435 F. Supp. 2d 36 (United States District Court for the District of Columbia). 2006. [Online]. Available: https://www.quimbee.com/cases/united-states-v-safavian

[72] United States v Branch, 970 F. 2d 1368 (United States Court of Appeals, 1992). 1992. [Online]. Available: https://casetext.com/case/us-v-branch-11

[73] People v Smith, 969 N.W.2d 548 (Michigan Court of Appeal). 2021. [Online]. Available: https://casetext.com/case/people-v-smith-7984

[74] Evidence Act 1995 (Cth) ss 146–7, 166–7, 170–2. 1995. [Online]. Available: https://classic.austlii.edu.au/au/legis/cth/consol_act/ea199580/

[75] S. Odgers, *Uniform Evidence Law*, 15th ed. Pyrmont, NSW, Australia: Thomson Reuters, 2020.

[76] S. Caruso, M. Legg, and J. Phoustanis, "The automation paradox in litigation: The inadequacy of procedure and evidence law to manage electronic evidence generated by the 'Internet of Things' in civil disputes," *Macquarie Law J.*, vol. 19, pp. 157–188, Nov. 2019.

[77] J. P. LaMonaca, "A break from reality: Modernizing authentication standards for digital video evidence in the era of deepfakes," *Amer. Univ. Law Rev.*, vol. 69, no. 6, pp. 1945–1988, 2020.

[78] N. Köbis, B. Doležalová, and I. Soraperra, "Fooled twice: People cannot detect deepfakes but think they can," *iScience*, vol. 24, no. 11, pp. 1–17, 2021. [Online]. Available: https://doi.org/10.1016/j.isci.2021.103364

[79] D. Harwell, "Top AI researchers race to detect 'deepfake' videos: 'We are outgunned,'" The Washington Post. Jun. 12, 2019. [Online]. Available: https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/

[80] K. Wade, S. Green, and R. Nash, "Can fabricated evidence induce false eyewitness testimony?" *Appl. Cogn. Psychol.*, vol. 24, pp. 899–908, Oct. 2010. [Online]. Available: https://doi.org/10.1002/acp.1607

[81] R. Chesney and D. Citron, "DeepFakes: A looming challenge for privacy, democracy, and national security," *California Law Rev.*, vol. 107, no. 6, pp. 1753–1820, 2019. [Online]. Available: https://doi.org/10.15779/Z38RV0D15J

[82] C. Nicholson, "A beginner's guide to generative AI." Wiki Pathmind. 2023. [Online]. Available: https://wiki.pathmind.com/generative-adversarial-network-gan

[83] J. Ruff, "The federal rules of evidence are prepared for deepfakes are you?" *Rev. Litigat.*, vol. 41, no. 1, pp. 103–126, 2021.

[84] C. Bhattacharyya et al., "Diffusion deepfakes," 2024, *arXiv.2404.01579*.

[85] Criminal code amendment (deepfake sexual material) bill 2024 (Cth). 2024. [Online]. Available: https://www.paulfletcher.com.au/parliamentary-speeches/second-reading-speech-criminal-code-amendment-deepfake-sexual-material-bill

[86] Disrupt explicit forged images and non-consensual edits act, S. 3696, 118th Congress. 2024. [Online]. Available: https://www.govinfo.gov/app/details/BILLS-118s3696is

[87] *Model Rules of Professional Conduct*, Amer. Bar Assoc., Chicago, IL, USA, 2020.

[88] *Legal Profession Uniform Law Australian Solicitors' Conduct Rules 2015, NSW, 2015*. [Online]. Available: https://legislation.nsw.gov.au/view/html/inforce/current/sl-2015-0244

[89] Chamberlain v Law Society of the Australian Capital Territory (1993) 43 FCR 148. 1993. [Online]. Available: https://jade.io/j/?a=outline&id=325954

[90] Hospital Products Ltd v United States surgical corporation (1984) 156 CLR 41, 1984. [Online]. Available: https://www.ato.gov.au/law/view/document?DocID=JUD/156CLR41/00005

[91] "Tech competence," LawSites, 2024. [Online]. Available: https://www.lawnext.com/tech-competence

[92] State ex rel Oklahoma Bar Association v Oliver, 369 P. 3d 1074 (Supreme Court of Oklahoma), 2016. [Online]. Available: https://law.justia.com/cases/oklahoma/supreme-court/2018/scbd-6728.html

[93] James v National Finance LLC, C.A. No.8931-VCL (Delaware Court of Chancery), 2014. [Online]. Available: https://law.justia.com/cases/delaware/court-of-chancery/2016/ca-8931-vcl.html

[94] M. Legg and F. Bell, "Artificial intelligence and the legal profession: Becoming the AI-enhanced lawyer," *Univ. Tasmania Law Rev.*, vol. 38, no. 2, pp. 34–59, 2019.

[95] A. McPeak, "Disruptive technology and the ethical lawyer," *Univ. Toledo Law Rev.*, vol. 50, pp. 457–476, May 2019.

[96] R. Jefferson and R. Pearce, "Not the end of lawyers, but the beginning," in *Leading Works in Legal Ethics*, J. Webb, Ed., London, U.K.: Routledge, 2023. [Online]. Available: https://doi-org.ezproxy.uow.edu.au/10.4324/9781003015093

[97] N. Yamane, "Artificial intelligence in the legal field and the indispensable human element legal ethics demands," *Georgetown J. Legal Ethics*, vol. 33, pp. 877–890, Sep. 2020.

[98] J. Littrich and K. Murray, *Lawyers in Australia*, 4th ed. Sydney, NSW, Australia: Federation Press, 2019.

[99] CRS Ltd v D'Arcy [1999] NSWCA 216, 1999. [Online]. Available: https://jade.io/article/123783

[100] R. Pfefferkorn, "Deepfakes in the courtroom," *Boston Univ. Public Interest Law J.*, vol. 29, no. 2, pp. 245–276, 2020.

[101] Mata v Avianca, Inc., 22-cv-1461 (U.S. District Court, S.D. New York, 2023), 2023. [Online]. Available: https://casetext.com/case/mata-v-avianca-inc-2

[102] M. Legg and V. McNamara, "AI is creating fake legal cases and making its way into real courtrooms, with disastrous results." The Conversation. Mar. 13, 2024. [Online]. Available: https://theconversation.com/ai-is-creating-fake-legal-cases-and-making-its-way-into-real-courtrooms-with-disastrous-results-225080

[103] M. F. Dahlstrom, "Using narratives and storytelling to communicate science with nonexpert audiences," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 4, pp. 13614–13620, Sep. 2014, doi: 10.1073/pnas.1320645111.

YVONNE APOLO
School of Law
University of Wollongong Australia
Wollongong, NSW 2250, Australia
E-mail: yapolo@uow.edu.au

KATINA MICHAEL
School for the Future of Innovation in Society
Arizona State University
Tempe, AZ 85287 USA

School of Computing and Augmented Intelligence
Arizona State University
Tempe, AZ 85287 USA
E-mail: katina.michael@asu.edu

Yvonne Apolo is a Lecturer with the School of Law, University of Wollongong. She is currently working on projects that interrogate the intersections between law and emerging technologies. Her primary research areas include the interaction between privacy law, psychology and emerging technologies, as well as the incremental development of tort law. She is an award-winning tertiary educator (OCTAL, 2019), and a member of the Legal Intersections Research Centre (LIRC), Wollongong Academy for Teaching and Learning Excellence (WATTLE) and an Associate Editor of the IEEE TRANSACTIONS ON TECHNOLOGY AND SOCIETY.

Katina Michael (Senior Member, IEEE) received the master's degree in transnational crime prevention from the School of Law, University of Wollongong in 2009. She is a Joint Professor with the School for the Future of Innovation in Society and the School of Computing and Augmented Intelligence, Arizona State University. She is also the Director of the Society Policy Engineering Collective and a Senior Global Futures Scientist with the Global Futures Laboratory. She is the Founding Chair of the Master of Science (Public Interest Technology) with Arizona State University in 2020. She is an Honorary Professor with the University of Wollongong, where she was previously the Associate Dean International in the Faculty of Engineering and Information Sciences, and was tenured from 2002 to 2022. She was previously employed as a Senior Network Engineer with Nortel Networks and also worked as a Systems Analyst with Andersen Consulting and OTIS Elevator Company. She has taught courses in the fields of computing and information technology, law and business, humanities and social sciences, and biomedical engineering. In 2017, she was awarded the Brian M. O'Connell Distinguished Service Award from the Society for the Social Implications of Technology. She is the Founding Editor-in-Chief of the IEEE TRANSACTIONS ON TECHNOLOGY AND SOCIETY. Previously, she was the Editor-in-Chief of *IEEE Technology and Society Magazine* from 2012 to 2017.