The Silicon Sentinel's Dilemma: Navigating AI as a Boon and Threat in India's Cybersecurity, Intelligence, and Judicial Frameworks

Vidhan Dilip Gambhire
University Department of Information & Technology, University of Mumbai

Abstract- **Artificial Intelligence has emerged as a disruptive force within the cybersecurity ecosystem in India, with extensive implications for national security, police operations, and the judicial process. Projections for 2025 are that AI-driven systems would increasingly form an integral part of cyber-defense operations, digital investigations, and forensic analysis infrastructures. These systems, while endowing capabilities of predictive threat intelligence, automated anomaly detection, and bulk data analytics, also introduce formidable technical, legal, and ethical challenges simultaneously. This article assesses Artificial Intelligence as a dual-use technology that enhances cybersecurity and also empowers sophisticated threats through deepfake fraud, data poisoning, algorithmic bias, and automated misinformation campaigns. Particular focus is paid to institutional barriers facing Indian intelligence agencies with respect to data overload and synthetically generated content, and how the judiciary is being increasingly dependent on AI-assisted evidence. The study proposes governance-directed initiatives, including the 'human-in-the-loop' mandates for judicial AI, explainable security architectures, standardized operational procedures for police forces, and mandatory digital watermarking of AI-generated government records. It contends that if deployed without human oversight, transparency, and accountability, Artificial Intelligence threatens to undermine public faith, due process, and the legitimacy of India's legal and security apparatus.**

The role of Artificial Intelligence is dramatically transforming the cybersecurity environment in the Indian context, from national security, law enforcement, and the judiciary. It is expected to integrate artificial intelligence tools in cyber security, investigatory work, surveillance networks, and forensic analysis by 2025. These tools promise immense benefits, from predictive threat analysis, automated anomaly detection, and the ability to examine massive datasets, but they also create difficult technology, legal, and ethical challenges. This article considers artificial intelligence as a mixed technology a double-edged sword, which can be harnessed for better cybersecurity, but also for more sophisticated cyber threats such as deep fake fraud, poisoning, biased algorithms, and automated campaigns for spreading misinformation. It underscores the challenges faced by Indian intelligence organizations in the context of the problem of data deluge, challenges from synthetic content, and the growing reliance on artificial intelligence for judicial evidence. There are also proposals for technology-based solutions for better cyber governance, such as maintaining human oversight for judicial applications, developing trustworthy security systems, establishing standard law enforcement practices, and using digital watermarking for all documents produced using artificial intelligence for government purposes. The claim is that artificial intelligence, without human oversight, could compromise public confidence, due process, and the sanctity and integrity of law enforcement and judicial bodies in the Indian context.

I. INTRODUCTION

The current legal landscape, encompassing acts such as the Information Technology Act of 2000 and sections of the Indian Penal Code of 1860, offers a rudimentary foundation, yet it conspicuously lacks specific provisions to govern the complex ethical and operational dilemmas posed by advanced AI applications, particularly within law enforcement [4], [6]. This gap in legislation brings out the need for a holistic legal framework that encompasses transparency, accountability, and ethics, which requires cooperation among legal experts, policy framers, and technologists. The India AI Mission 2024-2025 is one such strategic leap, wherein the nation is trying to transform its cybersecurity from conventional, reactive defense mechanisms to an AI-augmented security paradigm. This will achieve the use of artificial intelligence in predictive threat intelligence, automated anomaly detection, and adaptive cyber-physical system defenses, changing the very cores of operational dynamics in national security. However, this ambitious undertaking simultaneously introduces new vulnerabilities, as the very AI systems designed to protect could become

targets for sophisticated attacks, necessitating a robust and adaptive regulatory environment [8].

The effective realization of such an environment requires a critical evaluation of international frameworks, including the Algorithmic Accountability Act in the United States and the AI Act in the European Union, to guide the formulation of India's prospective legislative responses to algorithmic bias and data privacy concerns. These global precedents provide important insights for the establishment of comprehensive procedures to manage algorithmic inaccuracies and to enhance data privacy within AI governance, which are essential for fortifying India's legal regime and safeguarding individual rights. Moreover, the absence of specific AI legislation in India, despite the existence of broader policy frameworks, highlights a significant legislative gap that could impede the responsible development and deployment of AI technologies [9].

Therefore, India should adopt a contextual strategy that balances security, innovation, and civil liberties within its unique socio-legal environment instead of importing wholesale from foreign models. This involves proactive development of legal provisions to take care of algorithmic accountability and data privacy, for which the existing legislation, like the Information Technology Act of 2000, clearly falls short in considering the unique complexities brought up by AI. Current regulatory instruments such as the Telecommunications Act (2023) and the Digital Personal Data Protection Act (2023) mainly focus on cybersecurity aspects and data breaches, thereby presenting a significant regulatory vacuum relating to incidents concerning AI-unique operational events like performance degradation and algorithmic bias. The associated regulatory vacuum emphasizes the critical demand for a comprehensive AI-specific legal framework which would reduce the unique risks involved with the AI systems, especially in critical sectors like that of telecommunications. The ensuing governance gap is essentially manifested in a piecemeal and reactive oversight regime which falls short in addressing the AI-specific vulnerabilities like model drift, obscure decision-making, and algorithmic bias appearing increasingly in vital network operations. This calls for a comprehensive and forward-looking strategy in AI for achieving AI sovereignty and ensuring responsible AI development in India.

II. THE EVOLVING LANDSCAPE OF AI IN INDIA

The IndiaAI Mission (2024-2025) exemplifies a strategic pivot towards integrating advanced AI capabilities into the nation's critical infrastructure and defense mechanisms, moving beyond rudimentary cyber defense protocols to proactive, AI-driven threat intelligence [13]. This initiative seeks to harness AI for predictive analytics, automated anomaly detection, and adaptive responses to sophisticated cyber threats, fundamentally transforming national security paradigms [14]. This strategic adoption of AI, however, opens up a whole new dimension of vulnerabilities, where the very systems developed for protection may themselves become a target for sophisticated adversarial AI attacks, hence requiring strong and adaptive regulatory and technological defenses. These safeguards should be multi-faceted in nature, harnessing lifecycle governance, secure procurement, and strict oversight to ensure sovereign control over AI design and deployment. Furthermore, India's approach to artificial intelligence governance follows a different approach from that of other nations, one that is self-reliant and focused on economic growth within its unique socio-cultural context, rather than the generalized international regulatory frameworks. This would involve the development of regulatory models that specifically address the challenges of a large country with wide rural expanses and significant social issues, which thus reduces market disruption and preserves ethical oversight.

This strategic trajectory aligns with India's broader national AI strategy, which aims to position the nation as a global leader in AI while ensuring ethical and inclusive deployment, as articulated in comprehensive policy documents [18], [19], [20]. The aspiration to become a global AI superpower is further evidenced by India's increased interest in military AI, although limitations like the lack of a structured national roadmap, insufficient government investment, and quality AI talent currently impede this ambition [21]. Despite those challenges, the country is still dedicated to further development of its artificial intelligence ecosystem, which is illustrated by such projects as the IndiaAI Mission. It will involve the building up of talent and infrastructure necessary for future leadership of the country in AI. It also includes developing and deploying state-of-the-art AI systems, referred to as a "Sentinel Net,"

intended to bolster India's digital defenses and enhance capabilities for threat discovery, predictive analytics, and automated incident management protocols-focusing especially on critical infrastructure sectors.

This paradigm shift towards AI-driven security necessitates concurrent advancements in explainable AI and robust adversarial AI defenses to ensure the trustworthiness and resilience of these critical systems against evolving threats [23]. This emphasis on sovereign AI, however, requires careful consideration of the foundational compute infrastructure and energy demands, similar to strategies seen in other nations developing robust AI capabilities [16]. Furthermore, this pursuit of sovereign AI, marked by significant investments in national AI infrastructures like AI Factories, aims to secure local control over critical AI resources such as data, compute, and models, fostering innovation and economic competitiveness while mitigating technological dependence [24]. This localization effort is critical to resolve constraints unique to India-multilingual data, heterogeneous regulatory environments, and distinctive infrastructure deficiencies-which differ considerably from challenges found in Western AI development. The strategic implication for India, therefore, is to create an AI ecosystem that is technologically leading edge yet inherently robust and relevant to its unique geopolitical and societal imperatives.

III. FROM TRADITIONAL CYBER-DEFENSE TO AI-AUGMENTED SECURITY

This evolution necessitates a departure from static, signature-based defense mechanisms towards dynamic, AI-powered systems capable of real-time anomaly detection and adaptive threat response, crucial for mitigating sophisticated cyber threats [23]. This transition leverages machine learning algorithms to identify novel attack vectors and predict potential vulnerabilities before they are exploited, thereby enhancing the overall security posture [27]. Moreover, the integration of AI allows for the automation of routine security tasks, freeing human analysts to focus on more complex strategic challenges and threat intelligence [28]. AI-driven security measures, encompassing machine learning, deep learning, and predictive analytics, enable systems to proactively identify, mitigate, and respond to potential security vulnerabilities and attacks [29].

This shift transforms traditional incident response into a predictive and proactive defense, wherein AI models continuously learn from vast datasets to identify subtle indicators of compromise and orchestrate automated countermeasures [30]. This advanced capability drastically reduces reaction times, transforming cybersecurity from a reactive process into a proactive and anticipatory defense strategy [23]. This capability also enables scalable cybersecurity, wherein the management of increasingly volumes of data can be performed without a corresponding increase in human resources required to that end. The proactive capabilities accorded by AI in cybersecurity are particularly imperative given the increasing sophistication and proliferation of cyberattacks to date, underscoring the need for India to rapidly pursue indigenous AI development lest it incur a major national security deficit.

Such developments are in line with international trends, within which top nations such as the US are investing in significant AI infrastructure, talent development, and strategic research toward increasing their pool of AI-enabled cybersecurity capabilities. This transition from reactive to proactive defense strategies fundamentally changes how organizations approach cybersecurity, enabling the development of AI systems that detect subtle and complex threats with speed and accuracy far beyond human capabilities [31], [35]. This transition from manual, intensive cybersecurity practices to AI-enabled automation allows the scanning of large-scale datasets at unprecedented speeds, where complex patterns and their correlations can be identified to pinpoint advanced cyber threats that often evade traditional rule-based systems of detection. Therefore, AI cybersecurity solutions are vital in reinforcing defenses through proactive threat detection, predictive analytics, automated incident response, behavioral biometrics, and threat intelligence analysis.

IV. TECHNICAL CHALLENGES OF AI INTEGRATION

While considering artificial intelligence in cybersecurity, its potentials are inevitably entwined with a new raft of technically complex challenges that involve data integrity and explainability of algorithmic decisions. Key issues involve poisoning attacks on the training data AI models use, and the

so-called "black box" problem, where security decisions created by AI systems are hardly interpretable, making incident response and audit processes difficult. AI-assisted cybersecurity can therefore be effective only in conditions where the data being processed by the relevant algorithms is of good quality and trustworthy, which again makes the system susceptible to manipulation. Perhaps one specific point of recent vulnerability involves data poisoning-a sophisticated adversarial technique that involves the intentional introduction of corrupted or misleading data into machine learning model training datasets, with serious implications for integrity and operational behavior. Through this subversive tactic, models may incorrectly classify, create false positives or false negatives, or even establish covert backdoors that adversaries can use to leverage in future operations. The poisoning can degrade model performance, introduce biases, or otherwise allow targeted attacks by way of making the AI system unreliable or even exploitable. This is particularly concerning in critical Indian infrastructure networks, where compromised AI models could lead to widespread system failures or enable malicious entities to evade detection [27], [41]. Moreover, biased training data can lead to discriminatory outcomes in cyber threat detection and mitigation, further exacerbating vulnerabilities within these essential systems [38]. The implications extend beyond mere system failure, potentially enabling sophisticated cyber-espionage or even kinetic attacks if AI controls critical operational technology [42], [43]. This vulnerability underscores the urgent need for robust data validation protocols and adversarial training techniques to secure AI models against such sophisticated manipulation [30], [41]. Data Quality and Data Integrity Failure: The success of AI in critical infrastructure depends significantly on access to cleaned, high-quality datasets. However, their scarcity, often exacerbated by proprietary limitations or imbalanced distributions, can dramatically reduce model accuracy and generalization in live deployments [44]. Integration with Legacy Infrastructure: The difficulty of integrating high-end AI systems with India's diverse and often-antiquated legacy infrastructure further exacerbates these data-related challenges, often leading to interoperability issues, data format incompatibility, and heightened vulnerability to adversarial perturbations. These attacks aim to degrade the model's performance, introduce specific biases, or facilitate backdoor access, thereby corrupting the AI's decision-making process at a foundational level [46], [47]. For instance, adversaries can inject malicious data during training to degrade the performance of AI models, leading to incorrect predictions and potential financial or reputational damage [48].

Security of the AI Itself Beyond the integrity of its training data, the AI model itself can become a direct target for exploitation through adversarial attacks, where subtle perturbations to input data can cause misclassification, or through model exploitation techniques that extract sensitive information or compromise its decision-making logic [49], [50]. This type of manipulation is particularly insidious because it often remains undetected during standard audits, allowing corrupted foundations to persist within AI performance [51]. Furthermore, attackers can exploit zero-day vulnerabilities in AI systems to gain unauthorized access or disrupt operations, taking advantage of previously unknown weaknesses before patches are available [52].

A. Model Update Management and Version Control Challenges in Dynamic Environments:

Another vulnerability involves the intrinsic complexity brought on both by the frequent updating of AI models and by the associated rigorous version control in a dynamic operational environment. Security flaws or system stability might be inadvertently compromised in case of inconsistencies or un-verified changes. These problems are exacerbated by the "black box" problem: because many AI algorithms lack transparency, interpreting the steps of their decision-making processes, especially those which carry weighty security implications, becomes difficult. Due to a lack of explainability, sometimes termed the black-box phenomenon, notable challenges arise for forensic investigation, regulatory compliance, and accountability in significant cybersecurity events. This opacity prevents human operators from auditing AI decisions, detecting adversarial attacks, or determining whether a model has been compromised through data poisoning-representing a critical gap in the organization's security posture.

B. Real-Time Processing Constraints:

Furthermore, the computational demands related to real-time processing for advanced AI models in high-volume network environments can be overwhelming for existing infrastructure. This may result in delays or partial analyses, consequently compromising the

efficiency of threat detection and response mechanisms. The resulting opaqueness prevents not only the identification and mitigation of security vulnerabilities but also gives rise to some serious concerns about the trustworthiness and accountability of AI systems. This non-deterministic nature of AI models, where minor input variations can lead to unpredictable outcomes, complicates the task of ensuring security and reliability, especially when compounded by the rapid pace of AI development that often prioritizes innovation over comprehensive risk assessment [56]. Hence, the integrations of these models without robust security protocols and adversarial training leaves them susceptible to sophisticated attacks that exploit their inherent vulnerabilities [57], [58].

V. DATA POISONING IN INDIAN CRITICAL INFRASTRUCTURE

The illicit introduction of corrupted or manipulated data into AI training datasets poses a significant threat, especially to India's critical infrastructure, where compromised data can subtly alter model behaviors to facilitate undetectable cyber intrusions or disruptions. Data poisoning attacks may be in the form of "white-box" attacks, where adversaries possess full knowledge of the model and training data, or "black-box" attacks, which depend on using various sophisticated techniques to deduce model vulnerabilities without direct access to the model. These can compromise the integrity of crucial operational systems, leading to the misclassification of threats or the intentional deviation of automated responses within sensitive networks [61]. These vulnerabilities are even more disturbing, given that most sophisticated AI models come with limited transparency that complicates the detection of system compromise [62]. The latent character of data poisoning is further exacerbated by the fact that it can introduce backdoors, hence weakening security protocols, and enabling other more damaging cyber threats that could bring down critical services.

The impact of such attacks is amplified when considering the rapid deployment of inadequately tested AI products into business-critical applications [64]. Moreover, the expansive attack surface created by compound AI inference pipelines, which integrate multiple large language models, significantly increases the risk of successful data poisoning across interconnected systems [65]. This sophisticated manipulation can bypass fraud detection systems, distort algorithmic trading platforms to generate illicit profits, or cause significant market disruptions [66]. This expands the attack surface for malicious actors who can leverage AI supply chain vulnerabilities to introduce tainted datasets at various stages of model development and deployment [53]. Such attacks are especially perilous for open-source AI models, where the lack of rigorous provenance tracking can enable the widespread dissemination of poisoned models, as demonstrated by instances like "PoisonGPT" [67].

The objective of these attacks is to manipulate an AI model's decision-making by contaminating a substantial portion of its training dataset, potentially inserting hidden triggers that activate unpredictable behavior [68], [69]. This effectively renders AI systems susceptible to adversarial manipulation, allowing attackers to induce specific misclassifications or create backdoor access points that can be triggered at will [70]. This makes defense challenging, as the malicious data is often indistinguishable from legitimate input, necessitating advanced detection mechanisms and robust data validation protocols [71]. For instance, an attacker could control merely 0.01% of a dataset for a cost as low as US$60 to poison it, influencing AI models to provide biased responses or facilitate access for malicious actors [57], [72]. This underscores the urgent need for robust defense mechanisms, including stringent evaluation of data sources and continuous monitoring, to maintain the reliability of AI systems, particularly given that pre-trained models can propagate vulnerabilities across numerous integrated applications [48], [70], [73].

VI. THE 'BLACK BOX' PROBLEM: EXPLAINABILITY IN AI-DRIVEN SECURITY.

As AI is inherently opaque and with the advent of closed models and complicated deep learning models, it is impossible for the security analysts to have an actual understanding of how an AI is making such critical decisions about threats and which reactions are automated. This is not just hampering the debugging of processes but is an issue with respect to proving that the AI is regulation compliant and, most importantly, if an AI is under the effect of an advanced adversarial attack. Also, if there is no explainability associated with AI decisions and those

decisions are harmful because of which there is a significant shutdown of AI-critical infrastructure, there arises a need to revamp the usage of AI-powered security infrastructure. The intrinsic opacity of these advanced algorithms, often referred to as the "black box" phenomenon, makes it exceedingly difficult to ascertain the exact mechanisms leading to a particular security decision, thus challenging accountability and trust in mission-critical environments [35], [74].

VII. THE DUAL-EDGED SWORD: AI'S BOON AND THREAT

This section will examine the paradox of AI, its ability to improve cybersecurity on a large scale on one hand, but also create new, sophisticated threats on the other. This will be explored using the perspectives of the government, law enforcement, as well as citizens, and will examine how this subject is explored in relation to the use of the “Sentinel Net,” which provides surveillance and protection for the government, but also faces the increasingly prominent threat of “State-Sponsored AI Espionage.” Additionally, we will examine how AI provides protection for citizens in the form of “Phishing Guards” against the increasing threat of “Hyper Personalized Regional Language Scams.” This exploration seeks to understand the line between utilizing the capabilities of AI for increased security measures, but also ensuring it is not misused, particularly in the increasingly complex California Cybersecurity environment.

Additionally, as more industries become integrated with AI, the need for a comprehensive comprehension of relevant legal and ethical constructs, such as data privacy laws and algorithmic accountably, is imperative, especially within the context of the Indian cybersecurity environment, which lacks actionable details on specific implementations for AI technologies. The lack of established standards or norms for governing AI technology in India further complicates these issues, raising concerns about potential violations of civil rights as AI findings continue to be characterized by their opaque and incomprehensible nature [76]. Moreover, the absence of clear legislative frameworks for AI creation and deployment further exacerbates these concerns, leaving a void in addressing issues such as biased outputs and security vulnerabilities [58].

The Information Technology Act of 2000, for instance, was not designed to address the complexities of algorithmic accountability or data privacy in the context of advanced AI, leaving critical gaps in protecting individual rights [3]. These gaps highlight the urgent need for a comprehensive legal framework in India that specifically addresses AI governance, drawing insights from global models like the Algorithmic Accountability Act in the US and the EU's AI Act, which offer robust guidelines for managing algorithmic bias, ensuring data privacy, and enhancing transparency [3], [8]. Such a framework would not only safeguard individual liberties but also foster responsible innovation, ensuring that India's rapid AI adoption aligns with its commitment to a secure and equitable digital future [10], [18]. An architecture such as this would do much to safeguard individual rights; it would also provide a way to promote responsible innovation in the field, such that the Indian AI Gold Rush gets off to a good start down the right track towards a safe and equitable digital future. As things stand, however, the Indian judicial framework finds itself in a bit of a mess, without a rule book to speak to the new threats and opportunities posed by AI—primarily in the realm of data privacy and the mitigation of algorithmic bias. The lack of a rule book to guide the development of AI means that a legal framework to address the topic in broad terms—fusing concepts such as accountability, ethics, and transparency—must now take a prominent role.

However, existing digital laws in India are largely focused on cybersecurity and data-related breaches. This large gap becomes apparent when AI faces practical problems, such as when it slows down or is discriminatory. This lack of regulation highlights that India needs urgent AI legislation that covers these issues, rather than relying on a cybersecurity law that is too broad a parameter. This law would need to factor in India’s particular society while incorporating an international framework that is based on AI ethics.

The need for an agile framework becomes an essential aspect for India to keep the pace of rapid technological advancements and develop robust legal and ethical framework provisions for the developing Indian AI scenario. Taking into consideration the various complexities present, the regulatory framework for India needs to walk the line between the need for innovation and the ability to enforce stiff

measures perhaps through the concept of identifying high-risk categories of AI systems and developing strict ethical norms for technology advancement without-market-disrupting measures.

VIII. AI FOR GOVERNMENT AND LAW ENFORCEMENT AGENCIES

AI offers governmental and law enforcement agencies unprecedented capabilities for predictive policing, intelligent surveillance, and enhanced threat detection, transforming traditional reactive security paradigms into proactive defense strategies. This shift is evident in the deployment of AI-powered analytics for real-time monitoring of critical infrastructure and the sophisticated identification of anomalous patterns indicative of cyber threats or criminal activity [78]. These state-of-the-art tools enable us to sieve through big data in relatively less time, accelerating responses and making resource utilization more effective in national security-related work. But in the absence of a well-defined, unified national AI strategy and with policymaking scattered across India, we may miss these benefits and end up undermining the country's AI sovereignty. Such fragmentation not only delays the establishment of a robust regulatory framework but also creates critical gaps in addressing the ethical and social implications of AI deployed by the state machinery.

Furthermore, deployment of AI by the government requires rigorous oversight to check algorithmic bias and maintain non-discrimination and transparency, particularly in sensitive domains like law and order and intelligence operations. The integration of AI into government operations, while promising significant advancements, must therefore be meticulously managed to uphold democratic values and prevent potential abuses of power [79]. This meticulous management requires the establishment of a robust AI governance framework that includes clear ethical guidelines, accountability mechanisms, and independent auditing bodies to scrutinize AI algorithms used in public service [80]. A comprehensive and properly organized national AI strategy is a crucial requirement to guide responsible AI development and use in government domains. This increases public confidence, ensuring that scientific advancement is driven by public welfare and national security interests. Technology for the government is not merely about innovative solutions; it can enhance decision-making through real-time insights and analysis, which shape policies related to diverse domains, including city planning and responding to natural disasters. This has been coupled with ‘Sentinel Net’ strategic approaches set forth by AI technology. This inclusive cyber-protection strategy aims to enhance India's cyber environment through better risk forecasts and automated responses to cyber events. These systems leverage machine learning algorithms to continuously monitor network traffic, identify anomalous behavior, and respond to cyber threats with minimal human intervention, thereby augmenting the capabilities of human cybersecurity analysts [81]. Conversely, this extensive monitoring capability introduces significant concerns regarding potential misuse for mass surveillance, raising complex questions about privacy, civil liberties, and the scope of governmental oversight in a democratic society [82]. Moreover, the rise of state-sponsored AI espionage presents a formidable counter-threat, where hostile state actors leverage sophisticated AI to infiltrate critical infrastructure, exfiltrate sensitive data, and sow discord through information warfare [83]. This dual-edged nature necessitates a careful examination of AI's implementation, balancing its defensive potential against the risks of offensive deployment and algorithmic exploitation [84]. With the growing trend of using AI in governance, the need for rigorous legal and ethical policies is becoming imperative in mitigating threats. Such policies ensure that actions conducted by AI remain in check and in line with universal rights. This requires an effective tool of policymaking and new frameworks that define boundaries and ethics in using AI in the governance system.

IX. AI FOR CITIZENS: SAFEGUARDS AND RISKS

AI systems improve personal cybersecurity for laypersons with the help of AI-powered phishing protection software and intelligent spam filters. They read emails in a matter of seconds, identify harmful links, and recognize suspicious activity. But with this benefit comes a potential problem—the emergence of highly personalized and regional-themed scams using AI technologies to compose emails sounding as natural as possible. By using highly advanced natural language processing techniques and generative AI technologies, scammers can mimic actual communication styles so adeptly that it becomes virtually impossible to distinguish them from genuine emails.

This results in a highly potent form of a social engineering attack. The increasing cleverness of the scams developed using AI is why we need more intelligent tools that can identify and protect against such culture-specific threats developed using AI. Advanced warning tools must be able to use context and the most current threat intelligence to enable the tools to keep up with the evolving world of using AI for scams, helping protect individuals as we become increasingly active online. This is why we need to ensure we can adjust our defensive AI measures to protect against new techniques. Moreover, the integration of AI into public administration necessitates a delicate balance between leveraging its transformative potential and safeguarding against its inherent risks, including algorithmic biases and data privacy breaches, which demand robust regulatory oversight [85], [89], [90].

Effective AI governance, therefore, demands a collaborative approach involving government, industry, and civil society to develop comprehensive policies that address these multifaceted challenges while fostering innovation [91].

X. INSTITUTIONAL CHALLENGES AND VULNERABILITIES

The expanding reliance on AI across governmental and private sectors introduces significant institutional challenges, particularly concerning the volume and veracity of intelligence data. Intelligence agencies, such as the Intelligence Bureau and the Research and Analysis Wing, now face an unprecedented crisis of "Data Haystacks," where the sheer volume of information collected overwhelms human analytical capabilities, making it difficult to discern actionable intelligence from extraneous noise. This deluge is further complicated by the proliferation of deepfake misinformation, which sophisticated generative AI models can produce to intentionally mislead intelligence analysts and compromise national security through fabricated or distorted intelligence [92], [93]. This intentional obfuscation of facts, often indistinguishable from genuine data, poses a severe threat to accurate threat assessment and strategic decision-making, necessitating advanced AI-driven tools for veracity verification and anomaly detection [41]. These challenges underscore the urgent need for AI systems specifically designed to filter, prioritize, and validate intelligence inputs, thereby reducing the cognitive load on human analysts and enhancing the integrity of intelligence assessments [90]. However, the reliance on AI-driven mechanisms for intelligence analysis also introduces new vulnerabilities, particularly concerning the potential for algorithmic biases and the "black box" problem in AI decision-making, which can lead to flawed or inexplicable intelligence outcomes [90], [94]. Further complicating matters, the rapid adoption of AI across various sectors in India, while promising, also highlights a significant lag in the corresponding legal and regulatory frameworks necessary to govern its deployment effectively [95].

The uncertain environment spawned by the lack of clear legal frameworks regarding the criminal liability of AI as well as ethical considerations is a high-risk zone for innovation as well as national security. Let’s start with deepfakes. They are not mere data noise. They affect the very essence of intelligence efforts because they create AI-generated content with a plausible appearance of truth, upsetting such assessments. The dual-state nature of AI as a possible cyber danger as well as a means of automating forensics increases such challenges and brings a dire need for a reconsideration of the criminal jurisprudence of computational forensics integrity by the Indian system.

The increasing uses of AI technology by criminals to hide their identities and evade detection further obscure law enforcement’s capacity to trace the source of the attack and hold the responsible individuals liable, making the criminal responsibility of AI technology rather ambiguous. There appears to be a substantial gap in the existing legal framework, as core acts such as the Information Technology Act of 2000 and the Indian Penal Code of 1860 are inadequate to deal with the subtleties of using AI technology for criminal purposes and being held liable for the same. The need arises for the involvement of the High Courts and the Supreme Court of India to ensure that the use of AI technology remains aligned with the ever-growing legal parameters, particularly concerning the authentication of evidence produced by AI technology and the attribution of human intent by autonomous systems.

XI. INTELLIGENCE AGENCIES: DATA HAYSTACKS AND DEEPFAKE MISINFORMATION

The sheer volume of digital information, exacerbated by AI-generated content, creates a "data haystack" that overwhelms human analysts, making it exceedingly difficult to extract pertinent intelligence from the noise [98]. This challenge is compounded by the increasing sophistication of deepfake technology, which generates convincing but entirely fabricated audio, video, and textual content that can severely compromise intelligence integrity by introducing deliberately misleading information into these vast data sets [96]. Such AI-driven misinformation campaigns challenge traditional cybersecurity methods by blurring the lines between authentic and manipulated data, demanding a paradigm shift towards AI-enhanced detection and defense mechanisms [99]. The proliferation of AI-generated content also introduces new complexities for digital forensics, as distinguishing between genuine and synthetically altered evidence becomes increasingly difficult, impacting the admissibility and reliability of digital evidence in legal proceedings. This technological advancement necessitates a robust legal framework capable of addressing the criminal accountability of AI systems and their operators, particularly given the challenges of attributing intent and responsibility to autonomous entities [77], [100]. The traditional notions of mens rea and actus reus under Indian criminal law come into serious conceptual conflict when you attempt to extend these notions to AI.

This is because AIs lack human mental faculties and entail non-human physical actions that can't be specifically held responsible for their negative consequences. The issue at hand isn't addressed by current legislation like the Bharatyia Nyay Sanhita, 2023, and so it is imperative that lawmakers come up with new legislation that can define responsibility for crimes committed by means of AI. Moreover, AIs can generate 'hallucinated' cases, leading to greater complications in courts and making it imperative that validation checks be placed on any AI-cognitive assessment. Intelligence agencies, therefore, require sophisticated AI tools not only to sift through the data haystack but also to accurately identify and counter deepfake misinformation campaigns that could otherwise manipulate public opinion or compromise national security [102]. The increasing availability of advanced generative AI tools, capable of creating hyper-realistic text, audio, and video, exacerbates these challenges, making it easier to produce convincing but fabricated content that can be used to influence public opinion or disrupt political processes [103], [104]. This calls for a multi-layered approach, combining AI-driven anomaly detection with robust human verification protocols, to safeguard national security intelligence from increasingly sophisticated disinformation tactics. The unreliability and potential bias of current AI detection tools further complicate efforts to identify AI-generated material, whether genuine or fabricated, demanding continuous research into more robust and equitable detection methodologies [105]. This rapid pace of development also throws our detection tools back into the fray, and thus we find ourselves in what can be likened to a constant game of tug-of-war between what is possible on the generative side of AI and our reaction methods on the defensive side of AI. Nor can we be careless in the area of ethics, especially in terms of privacy and the consequences of bias on particular sections of society.

XII. JUDICIAL SYSTEM: AI-ENHANCED EVIDENCE AND HALLUCINATED PRECEDENTS

The applicability of AI-informed evidence within Indian courts, such as CMOS image analysis powered by AI or AI-produced forensic documents, is set to be strictly regulated through legal means, most specifically "the Bharatiya Sakshya Bill," a new Evidence Act that effectively replaces the existing Indian Evidence Act. This particular worry is no mere triviality, given that many AI applications are effectively "black boxes," making it difficult to distinguish how a particular AI application could arrive at a particular conclusion that could then be used as evidence. Furthermore, there is also a phenomenon known as AI "hallucinations," where AI produces attractive but false data, potentially mistaken for actual facts or precedents. Such measures are crucial not only for maintaining the integrity of judicial processes but also for establishing public trust in AI's role within the legal system, particularly as AI tools are increasingly considered to assist with the adjudication of cases to address pending caseloads [106]. Moreover, the challenge of "Impostor Bias" [107] further complicates the acceptance of AI-generated content in legal contexts, as the distinction between human-

created and AI-generated evidence becomes increasingly difficult for human observers to discern. This cognitive bias can lead to the devaluation of authentic media or, conversely, the acceptance of fabricated evidence as genuine, underscoring the critical need for advanced forensic tools and rigorous judicial training to evaluate AI-generated media [107].

The critical role of data protection authorities in ensuring compliance with privacy regulations becomes paramount in this landscape, particularly concerning AI development and its potential impacts on fundamental rights [18]. The opacity of AI systems, often termed the "black box" phenomenon, makes it challenging for courts to examine the underlying factors leading to AI-generated testimony, thus diminishing its evidentiary value [108], [109]. The decision-making process of AI-operated entities that is shrouded in secrecy strikes at the very heart of due process and fair hearing rights. If the detailed rationale for the conclusions drawn on any evidence is not presented before all parties, then questions of fairness will arise. Thus, legal obstacles for the development, deployment, and governance of AI need to be overcome for its application in such a manner as not to infringe individual rights, make the playing field uneven, and erode public trust. Mastering such tricky terrain requires the law to move swiftly, in response to AI's particular features and the specific potential for abuse, yet open up the potential for legitimate applications of AI within criminal justice. Given these complexities, it is vital to develop comprehensive guidelines for the rigorous evaluation and benchmarking of AI systems intended for legal applications, drawing parallels with established standards in other critical fields like biometric evaluation [113]. Therefore, mitigating inherent biases within AI algorithms and ensuring their explainability becomes a paramount concern to prevent discriminatory outcomes and maintain judicial integrity [101], [114], [115].

XIII. SCENARIO ANALYSIS IN INDIA

This section explores hypothetical yet plausible scenarios for India in 2025, illustrating the multifaceted impact of AI across cybersecurity, law enforcement, and judicial integrity. These scenarios underscore the urgent need for proactive policy development and robust technological safeguards to harness AI's benefits while mitigating its inherent risks within the Indian context. The absence of clear regulatory frameworks governing AI's application in India currently poses significant challenges to addressing these emerging issues, potentially allowing AI misuse to infringe upon fundamental civil rights [98]. The rapid proliferation of AI technologies, coupled with insufficient legal and ethical guidelines, creates a fertile ground for novel forms of cybercrime and sophisticated challenges to evidentiary standards in legal proceedings [11], [116]. This regulatory vacuum also allows for the exploitation of AI vulnerabilities, leading to scenarios that directly threaten national security and public trust [5], [76]. Consequently, the Indian legal system must adapt swiftly to these technological advancements, formulating clear regulations that define the admissibility of AI-generated evidence and establish accountability for AI-driven decisions [11], [117]. Moreover, the ongoing debate regarding the appropriate legal frameworks for AI liability, particularly in cases of autonomous AI systems, highlights the urgent need for India to develop comprehensive legislative responses that can withstand future technological advancements [108], [118].

Examining the role of AI-powered CCTV and the freedoms of citizens. Just envision a scenario where top-of-the-line, AI-powered CCTV systems are installed in the guise of safety and prevention of crimes when, in fact, these systems are stealthily squeezing the freedoms of privacy and movement out of citizens. The balance between securing citizens and upholding these freedoms turns out to be a walk on a thin rope. Algorithmic discrimination, which may potentially discriminate against particular sections of citizens, is an added twist that unfolds.

Such highly effective surveillance technology promises to ensure greater security for citizens but does so at the cost of some challenging ethical considerations concerning the use of one’s data and the possibility that one’s surveillance data could potentially be abused. If this technology were to lack effective regulation, there could potentially be the issues of mission creep and the unauthorized use of data.

AI-Powered Call Analysis in Cyber Fraud: examining actual telecom transmissions in the new light of AI. The text explores both sides of a reality. One, AI could be the huge booster of law

enforcement efforts to actually identify calls as fraud when it takes place. The flip side, however, is the immense possibilities of threats to the right to privacy. The key is getting it right – protecting genuine threats with the same suffocating it not to infringe the right to privacy. Moreover, by introducing AI into the context of calls being analyzed, several practical issues surface with regard to handling calls, namely avoiding leaks and staying within legal frameworks when it comes to data protection. As far as the law is concerned, using AI to analyze calls, including those that entail eavesdropping and storing communications, come under the scanner when it comes to applicable Indian laws and rules when the upcoming Digital Personal Data Protection Act comes into play.

Local Issue: Plagiarism Lawsuit PK This particular situation focuses on the issue that may arise when, for instance, a deep fake, which is a very convincing version of a video or audio clip, is used to make a complaint against any random citizen. It is pertinent to mention in this context that it seems to be quite challenging for the existing methods for validating digital evidence to include it in their capabilities, and there is also the question of being mistakenly implicated.

"The increasing capabilities in the area of deepfakes require a response that is not only technological, but must also entail a rethinking of the way in which evidence online is treated in India, including the burden of proof and the character of credible proof in the justice system."

The AI 'probable manipulation' score
When the score undergoes incorporation within the judicial system, the question of its validity and the ability of human beings to review its interpretations also arises. This becomes more significant as the ability of the AI model to hallucinate or present erroneous scores also arises. The police might feel reluctant to trust the decision of the AI model, and thereby, the human review approach becomes more necessary.

The uncertainty and often opaque nature of the findings of AI systems has raised significant international concern, making it even more challenging to integrate AI findings into an existing legal and investigative framework. The problem of the so-called 'black box,' where the rationale of the AI decision isn't transparent, further adds to the challenges of basing an assessment of AI on legal standards, as it's quite challenging to determine the criteria for a given 'manipulation score.'AI in Missing Persons & Criminal Tracking This scenario evaluates the deployment of AI systems to enhance the efficacy of missing persons searches and criminal tracking by analyzing vast datasets, raising critical questions about individual privacy and potential for algorithmic bias in profiling.

The unprecedented scale of data processing involved necessitates robust ethical guidelines and stringent data protection frameworks to prevent the misuse of personal information and safeguard civil liberties [120]. Moreover, the application of AI in such sensitive domains demands rigorous validation of algorithmic fairness to mitigate the risk of biased outcomes, which could disproportionately affect marginalized communities [121]. The integration of AI in these critical areas also necessitates a thorough examination of data anonymization techniques and consent mechanisms to ensure compliance with privacy regulations [18]. Faster identification of individuals through facial recognition and predictive analytics tools offers significant operational advantages to law enforcement agencies [122]. Risk of Wrong person picked up for verification The potential for erroneous identification through AI systems, as evidenced by cases where individuals have been wrongfully arrested based on faulty facial recognition [123], underscores the critical need for human oversight and verification protocols in AI-driven law enforcement applications.

XIV. DEEPFAKE VOICE-CLONE SCAM TARGETING STATE TREASURY

This hypothetical scenario illustrates the profound vulnerability of critical financial infrastructure to sophisticated AI-powered impersonation tactics, necessitating advanced biometric verification and anomaly detection systems. In this scenario, a state-of-the-art voice-cloning AI, leveraging publicly available audio samples of a high-ranking treasury official, successfully bypasses conventional authentication protocols to authorize fraudulent financial transfers. This insidious act exploits the increasing realism of generative AI, which has introduced transformative tools capable of producing complex, human-like audio that can deceive even trained ears and automated systems [124]. The

scenario highlights how even advanced AI surveillance and authentication systems can be compromised by equally sophisticated generative AI, posing significant threats to financial security [99]. The ability of generative AI to create highly realistic fake identities and fabricate sophisticated financial documents further complicates fraud detection, making it increasingly difficult for human reviewers to discern forged records from genuine ones [66]. Such incidents underscore the urgent need for robust multi-factor authentication systems and AI models specifically designed to detect AI-generated forgeries, thereby enhancing the resilience of financial institutions against evolving cyber threats [125], [126]. The proliferation of voice cloning capabilities, achievable from mere seconds of audio, fundamentally challenges established voice biometric authentication landscapes and has led to a surge in sophisticated scams [127], [128]. This technological advancement enables threat actors to mimic trusted individuals with unprecedented accuracy, leading to significant financial losses and erosion of trust in digital communication [129], [130]. These deepfake voice scams can exploit vulnerabilities in traditional authentication methods, enabling criminals to impersonate authorities or trusted individuals to commit financial fraud on a large scale [131], [132]. Such incidents have already been documented, with cases involving millions of dollars being defrauded through sophisticated voice-cloning operations targeting financial institutions and individuals alike [66], [133], [134]. For instance, a Hong Kong bank manager authorized a $35 million transfer based on an AI-cloned voice, demonstrating the devastating potential of such attacks [135]. This underscores the critical need for financial institutions to reassess the security of voice-based authentication systems and invest in advanced AI-driven detection mechanisms capable of discerning AI-generated speech from authentic human voices [66], [136]. Further, these AI-driven phishing attacks leverage natural language processing to craft emails and messages indistinguishable from legitimate communications, increasing their efficacy in extracting sensitive information [66]. The sophisticated nature of these attacks, often involving hyper-personalized content, necessitates a paradigm shift in cybersecurity strategies, moving beyond traditional signature-based detection to advanced behavioral analytics and AI-driven anomaly detection to identify subtle indicators of compromise [128]. The availability of malicious AI systems like FraudGPT and WormGPT in underground markets further exacerbates this threat, providing sophisticated tools for crafting highly convincing and tailored phishing messages [137]. The ease of access to such platforms, coupled with the increasing sophistication of voice cloning technologies, has made it simpler for even low-skilled threat actors to engage in advanced cybercrimes, posing a significant challenge to existing cybersecurity defenses [67]. Furthermore, the advanced capabilities of generative AI now enable attackers to impersonate multiple individuals in a single video conference, as demonstrated by a $25 million fraud where every participant was a deepfake, thereby circumventing established multi-factor authentication protocols [138]. India acces a significant challenge in mitigating these advanced threats, particularly given the rapid adoption of digital payment systems and the burgeoning digital economy, which provide a fertile ground for such sophisticated AI-driven scams [138], [139]. Solutions must therefore encompass not only technological advancements in AI-driven fraud detection but also widespread public awareness campaigns to educate citizens on the evolving tactics of AI-powered social engineering [30], [47], [139].

XV. AI-DRIVEN BREACH OF SMART CITY TRAFFIC GRID

With this in mind, the paper discusses the ways in which rogue AI could manipulate weak points in a smart-city connected network to create widespread disruption and safety hazards. A cyber-physical attack focused on using AI to manipulate traffic-management systems can bring traffic to a standstill, block emergency responders, or even spur multi-vehicle crashes-a new generation of urban conflict. It would be able to exploit poorly secured IoT devices for entry into the general city network. Increased proliferation of insecure smart sensors and unpatched network hubs within smart cities presents a huge attack surface for AI-driven exploits. The complex interplay of AI-powered systems with old infrastructure further exacerbates these gaps, increasing the likelihood of cascading failures triggered by a single well-orchestrated AI attack. Absence of universal standardized cybersecurity rules across diverse components of a city makes detection of, and defense against, sophisticated AI intrusion significantly more difficult. This emphasizes the importance of good AI governance, continuous security auditing of all AI-powered urban

systems, and an aware and adaptive security posture. Since smart-city assets are complex and connected, only an AI threat intelligence-driven and real-time anomaly detection-powered proactive approach will help in neutralizing such sophisticated threats well in advance.

The adversarial manipulation of AI systems within smart city grids, particularly through data poisoning, poses a critical threat by leading to misclassified traffic conditions and improper signal adjustments, ultimately causing significant gridlock and hindering emergency responses [142]. Furthermore, such attacks can extend beyond traffic management to disrupt critical infrastructure components such as energy distribution and water supply, exacerbating urban chaos and compromising public welfare [143].

In India smart cities like Bhubaneswar, Kakinada, and Pune, which heavily rely on interconnected IoT devices and AI for operational efficiency, are particularly vulnerable to these advanced cyber threats [144], [145], [146]. and also financial losses associated with remediation and reputational damage. Cities like financial captial of India Mumbai city face estimated cyber incident costs averaging $3.4 million, encompassing direct financial losses, reputational harm, and extensive recovery expenses [144] and Navi Mumbai now upgrading and developing its smart city infrastructure, further amplifying the need for robust AI-driven cybersecurity measures to protect against sophisticated attacks [144], [147]. Such vulnerabilities are not merely theoretical; the interconnected nature of smart city infrastructures means a breach in one system can compromise the entire network, necessitating robust encryption and secure communication protocols [144].

XVI. PROPOSED SOLUTIONS AND POLICY RECOMMENDATIONS

The need to address the set of complex issues that AI raises for the field of cybersecurity, the police, or the fairness of judicial proceedings requires not only technological solutions but a wide-ranging strategy incorporating technology and regulation. The following is a policy that could help address these issues: to ensure that all decisions made by AI in the judiciary must have a Human-in-the-Loop, and to ensure that all government records that are created by AI must feature Digital Watermarking.

Developing AI that is transparent and trustworthy is also important, as this will make it possible to analyze decisions that AI is making, and this will increase confidence in how AI systems are working in a legal or a governmental setting. Regulators should insist on transparency in law enforcement about what tools of AI are used, data input into these tools, and how these tools may affect society.

A strong legal framework embracing the strands of transparency, accountability, and ethics, developed after joint efforts of various stakeholders, is essential for the effective adoption of the benefits of AI, while preserving the integrity of the legal frameworks of India. Standard Operating Procedures (SOPs) for AI Use by LEA, Developing clear Standard Operating Procedures for law enforcement agencies in their use of AI is crucial to ensure adherence to ethical guidelines, legal frameworks, and human rights principles, especially given the rapid adoption of AI in policing [99].

These procedures should encompass data privacy protocols, bias mitigation strategies, and clearly defined scenarios for AI deployment, ensuring that AI tools augment, rather than supersede, human judgment and oversight in critical decision-making processes [99]. Regular, independent audits of AI systems for bias and conceptual drift are imperative to ensure their continued fairness and accuracy, particularly in applications related to law enforcement and judicial processes [149]. Also, these audits should be accompanied by checklists that entail information pertaining to the theory underlying crime prevention, the nature of crimes being focused on, the sources of the data used, the technology used, and measures of success. Also, these audits should provide summary reports.

Clear Citizen Appeal and Review Mechanism It is important to have a clear and transparent process through which individuals impacted by the decisions of artificial intelligence systems would have the ability to appeal and seek recourse. It needs to have human oversight in which the decisions of the artificial intelligence systems are verified and checked to ensure that they are not biased or come to incorrect decisions through the systems.

Dataset Sovereignty & Contextual Training To further enhance the reliability and fairness of AI systems, particularly in sensitive domains like law

enforcement, it is crucial to ensure that training datasets are not only comprehensive but also contextually relevant to India's diverse socio-cultural landscape, thus preventing the perpetuation of biases inherent in generalized global datasets [120]. This requires significant investment in collecting and annotating localized data, reflecting India's unique demographics and legal precedents [20]. Such localized datasets, carefully curated and regularly updated, will enable AI models to make more accurate and equitable decisions that resonate with the specificities of Indian society and legal interpretations [120]. Furthermore, strict regulations must govern the acquisition and use of AI systems by law enforcement, mandating auditable and accessible algorithms for independent evaluation to ensure compliance with fundamental rights and applicable laws [151].

XVII. HUMAN-IN-THE-LOOP REQUIREMENT FOR JUDICIAL AI

The integration of a "Human-in-the-Loop" model in judicial AI systems is not merely a best practice but a foundational necessity to ensure the integrity and accountability of legal decisions, particularly in complex Indian legal contexts. This approach mandates continuous human oversight and intervention at critical junctures of AI-driven judicial processes, mitigating risks such as AI "hallucinations" and biased outputs that could undermine justice [152]. This demands that human jurist experts evaluate AI-assisted evidence and verdicts judiciously, ensuring that these conform to existing jurisprudence and preventing AI systems from generating incorrect and unjustifiable precedents. Human intervention is imperative when it comes to detecting and addressing biases existing in AI-assisted data sources, preventing potential AI tools for creating unjustifiable and existing imbalances that discriminate systematically against marginalized groups through an unjust court system. This serves as an additional layer towards preventing the issue of black boxes that would allow assessment and interpretation of AI-assisted analytical solutions through interpretation prior to their application. The need for a human-based system would be imperative, especially if placed within the context of an India-based system that would require expertise when it comes to placing appropriate AI-assisted analytic perspectives appropriately.

The current attempts by the Supreme Court of India to transcribe arguments and trials would represent an emerging recognition of an appropriate system that must incorporate AI supported by jurist supervision that could enhance efficiency while maintaining accuracy. The system would represent an appropriate recognition that would emphasize AI-based tools that must complement jurist expertise that would otherwise require substitution, especially when addressing critical segments involving expertise related to matters of law. Implementing a human-in-the-loop system in legal AI specifically aims to enhance decision-making by combining AI's computational power with human contextual judgment [153]. Such an approach is particularly vital in the Indian legal system, characterized by its intricate jurisprudence and the profound impact of cultural and social factors on legal outcomes, requiring human experts to discern the subtleties that AI alone might miss [153]. Hence , the human element acts as a crucial check, preventing AI from solely relying on historical data, which might otherwise perpetuate systemic biases prevalent in past legal precedents [156].

XVIII. MANDATORY DIGITAL WATERMARKING FOR AI-GENERATED GOVERNMENT RECORDS.

The use of digital watermarking on all AI-produced government documents is an critical component in the preservation of the authenticity, integrity, and traceability of documents in a world where the role of AI in the administration of public documents will continue to grow. This method will enable these documents to be easily identifiable if they are produced or if they have undergone substantial enhancement by AI. This action will not only protect the documents from fraudulent manipulation but will also provide a mechanism to audit the effect of AI on the formulation of laws. This will enable the government to know the role AI plays in informing its laws. This digital watermarking also facilitates the verification of AI-generated content by allowing for independent review of its origin and any subsequent modifications, thereby bolstering transparency and accountability in governmental operations [158]. Such a system would also be instrumental in legal contexts, enabling courts to readily distinguish between human-authored and AI-generated evidence, which is especially pertinent given the challenges posed by AI-enhanced evidence in

judicial proceedings This proactive move enables the verification of the veracity of digital documents each and every time, reducing the possibility that misleading precedents or sham evidence sneaks through the courts. Digital watermarks, while machine-readable, should preferably be human-noticeable too, so we clearly see when content comes from AI. That transparency helps alleviate concerns about the fragility of AI watermarking for images, providing a robust and standard means to check for originality. Add to that state-of-the-art watermarking technology, which embeds information within AI-generated outputs in a very difficult-to-remove or tamper-with way, so even sophisticated attempts to alter content can be detected. It is this robust form of watermarking that enhances accountability by providing the ability to trace not just the specific model of AI but also the responsible parties in question should something go wrong or be used in a malicious manner. Moreover, this mandatory watermarking would provide a critical mechanism for attributing content, enabling future AI models to prioritize high-quality human-created data and thereby preserve the integrity of training datasets against potential contamination [165]. This approach aligns with international regulatory trends, such as the EU's AI Act and the US Executive Order 14110, which increasingly mandate provenance tools and machine-readable content markings to ensure accountability and transparency in AI-generated content [166]. India, with its ambitious IndiaAI Mission, stands to benefit significantly from such a framework, fostering public trust and safeguarding the integrity of its digital infrastructure [167]. Ai policy in India must therefore incorporate these robust technical and legal frameworks to proactively mitigate the risks associated with AI's rapid integration into critical governmental functions [167], [168].

XIX. CONCLUSION

Artificial Intelligence is quietly being integrated with the cybersecurity, police, and support justice systems of the country, India. It is pertinent to note that, according to this research, it is no longer about the optimization of technology with the advent of Artificial Intelligence, which actually influences the detection, inquiry, and evaluation of threats. As of 2025, decision-making in major sectors with regard to National Priority will be accomplished with the help of Artificial Intelligence.

AI improves cyber-security through quick threat detection, predictive analysis, and automated defenses. However, it also introduces challenges. Data poisoning, adversarial attacks, deepfake fraud, algorithmic opacity, and bias can impact both the dependability and validity of AI. In India, these challenges matter. Intelligence agencies simultaneosuly manage huge volumes of data, while courts depend on AI-enhanced digital evidence. The absence of independent legislation on AI can worsen these challenges. This forces institutions to fall back on existing legislation that was not intended to govern self-driving or semi-autonomous decision-making.

This study emphasizes that the risk is not in the use of AI, but in the uncritical faith placed in its output without any mechanism to validate it. When the computers function like “black boxes,” the integrity of investigation is compromised, and the process of justice itself is endangered. In courts, the admissibility of evidence not verifiable by humans can certainly jeopardize justice itself. Therefore, an AI future for India must be guided by governance first. Models with human-in-the-loop capabilities, explainable and auditable AI systems, standards of policing by the law enforcement agencies, and watermarking of all AI-produced government documents are necessary safeguards. This is to keep AI systems as an aid and not an authority.

Consequently, AI must be a controlled enabler of cybersecurity and justice, and never a replacement for human discretion. The appropriate application of the tool, within the framework of legal and ethical accountabilities and governance, will play an essential role in making sure that the tool does nothing to dilute the foundational values of the Indian Constitution.

REFERENCES

[1] T. Meghwal, “Emerging Challenges in Regulating Artificial Intelligence Under Cyber Security Laws in India,” Jan. 2025, doi: 10.2139/ssrn.5030258.

[2] S. Sahibpreet and D. Shikha, “Cybercrime and Computer Forensics in Epoch of Artificial Intelligence in India,” *arXiv (Cornell University)* , Dec. 2025, doi: 10.48550/arxiv.2512.15799.

[3] S. A.V. and N. D. N. -, "Legal Challenges of Artificial Intelligence in India's Cyber Law Framework: Examining Data Privacy and Algorithmic Accountability Via a Comparative Global Perspective," *International Journal For Multidisciplinary Research* , vol. 6, no. 6, Nov. 2024, doi: 10.36948/ijfmr.2024.v06i06.31347.

[4] A. Sharma, S. Sharma, S. D. Soni, P. Agrawal, P. K. Mishra, and G. Mourya, "Artificial Intelligence in the Indian Criminal Justice System: Advancements, Challenges, and Ethical Implications," *Journal of Lifestyle and SDGs Review* , vol. 5, no. 1, Jan. 2025, doi: 10.47172/2965-730x.sdgsreview.v5.n01.pe04877.

[5] N. Kumar, "From Surveillance to Sentencing: Evaluating AI's Role in Indian Criminal Justice," *SSRN Electronic Journal* , Jan. 2025, doi: 10.2139/ssrn.5391285.

[6] H. Sayyed, "Artificial intelligence and criminal liability in India: exploring legal implications and challenges," *Cogent Social Sciences* , vol. 10, no. 1, Apr. 2024, doi: 10.1080/23311886.2024.2343195.

[7] A. Agarwal and M. J. Nene, "Incorporating AI Incident Reporting into Telecommunications Law and Policy: Insights from India," 2025, doi: 10.48550/ARXIV.2509.09508.

[8] A. Lamba, P. N. Nayyar, and T. Tanwar, "Cybersecurity Laws and Privacy Protection in India," in *Advances in Social Science, Education and Humanities Research/Advances in social science, education and humanities research* , 2025, p. 22. doi: 10.2991/978-2-38476-426-6_3.

[9] N. Basu and R. Dave, "Comparative Analysis of Laws in AI," *Journal of Lifestyle and SDGs Review* , vol. 5, no. 3, Feb. 2025, doi: 10.47172/2965-730x.sdgsreview.v5.n03.pe05575.

[10] [10] P. Singh, "Regulating decentralized AI: Blockchain, dark web, and anonymity challenges, an Indian perspective," *International Journal of Law Justice and Jurisprudence* , vol. 5, no. 2, p. 220, Jul. 2025, doi: 10.22271/2790-0673.2025.v5.i2c.234.

[11] I. Biswal and R. Verma, "Artificial Intelligence in Legal Practice: Implications for the Legal Profession's Future," *International Journal for Research in Applied Science and Engineering Technology* , vol. 12, no. 4, p. 1974, Apr. 2024, doi: 10.22214/ijraset.2024.60199.

[12] A. Avinash, A. Peeyush, and N. M. J, "AI Regulation in Telecommunications: A Cross-Jurisdictional Legal Study," *arXiv (Cornell University)* , Nov. 2025, doi: 10.48550/arxiv.2511.22211.

[13] J. Vipra, "Towards AI sovereignty: The good, the bad, and the ugly of AI policy in India," *The African Journal of Information and Communication (AJIC)* , no. 35, p. 1, Jul. 2025, doi: 10.23962/ajic.i35.21263.

[14] K. Nilgiriwala *et al.* , "Navigating the Governance of Artificial Intelligence (AI) in Asian Nations: A Focus on India, Indonesia, Malaysia and the Philippines," *SSRN Electronic Journal* , Jan. 2024, doi: 10.2139/ssrn.4735279.

[15] S. Singh and S. Sengupta, "Sovereign AI: Rethinking Autonomy in the Age of Global Interdependence," *arXiv (Cornell University)* , Nov. 2025, doi: 10.48550/arxiv.2511.15734.

[16] S. B. Chetty *et al.* , "Sovereign AI for 6G: Towards the Future of AI-Native Networks," *arXiv (Cornell University)* , Sep. 2025, doi: 10.48550/arxiv.2509.06700.

[17] Y. Walter, "Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences," *Discover Artificial Intelligence* , vol. 4, no. 1, Feb. 2024, doi: 10.1007/s44163-024-00109-4.

[18] N. Joshi, "Emerging Challenges in Privacy Protection with Advancements in Artificial Intelligence," *International Journal of Law and Policy* , vol. 2, no. 4, p. 55, Apr. 2024, doi: 10.59022/ijlp.171.

[19] A. K. Agarwal and M. J. Nene, "A five-layer framework for AI governance: integrating regulation, standards, and certification," *Transforming Government People Process and Policy* , vol. 19, no. 3, p. 535, May 2025, doi: 10.1108/tg-03-2025-0065.

[20] S. Bansal and N. Jain, "A Comprehensive Study Assessing the Transformative Role of Artificial Intelligence in India\'s Governance Policy Framework," *International Journal for Research in Applied Science and Engineering Technology* , vol. 11, no. 7, p. 1748, Jul. 2023, doi: 10.22214/ijraset.2023.54973.

[21] S. Mohan, "Passive Ambitions, Active Limitations: Defence AI in India," 2024, p. 445. doi: 10.1007/978-3-031-58649-1_20.

[22] R. Chakrabarti and K. Sanyal, "Towards a 'Responsible AI': Can India Take the Lead?," *South Asia Economic Journal* , vol. 21, no. 1, p. 158, Mar. 2020, doi: 10.1177/1391561420908728.

[23] A. Shrivastav, "AI-Driven Cyber Defence for India's National Security," *Electronic Journal of Social & Strategic Studies* , vol. 6, no. 7, p. 139, Jan. 2025, doi: 10.47362/ejsss.2025.6608.

[24] P. G. Lopez *et al.* , "AI Factories: It's time to rethink the Cloud-HPC divide," 2025, doi: 10.48550/ARXIV.2509.12849.

[25] R. Raju, "From Models to Markets: Generative AI and Its Emerging Role in Indian Financial Services," *SSRN Electronic Journal* , Jan. 2025, doi: 10.2139/ssrn.5223947.

[26] M. G. Jacobides, S. Brusoni, and F. Candelon, "The Evolutionary Dynamics of the Artificial Intelligence Ecosystem," *Strategy Science* , vol. 6, no. 4, p. 412, Oct. 2021, doi: 10.1287/stsc.2021.0148.

[27] R. Arslan, M. Özseven, and M. M. Aydın, "Transforming European Cybersecurity: AI-Powered Threat Analysis, Quantum Age, Blockchain/Crypto Risks, and Regulatory Strategies," *International Journal Of Engineering & Applied Sciences* , vol. 17, no. 2, p. 81, Sep. 2025, doi: 10.24107/ijeas.1619600.

[28] Y. Fernandes and N. Abosata, "Analysing India's Cyber Warfare Readiness and Developing a Defence Strategy," *arXiv (Cornell University)* , Jun. 2024, doi: 10.48550/arxiv.2406.12568.

[29] S. Rangaraju, "SECURE BY INTELLIGENCE: ENHANCING PRODUCTS WITH AI-DRIVEN SECURITY MEASURES," *EPH - International Journal of Science And Engineering* , vol. 9, no. 3, p. 36, Dec. 2023, doi: 10.53555/ephijse.v9i3.212.

[30] Sweety, "The Rise of AI-Powered Cybersecurity Threats and the Evolution of Defense Mechanisms," *International Journal for Research in Applied Science and Engineering Technology* , vol. 13, no. 5, p. 7215, May 2025, doi: 10.22214/ijraset.2025.71745.

[31] M. Roshanaei, M. R. Khan, and N. N. Sylvester, "Enhancing Cybersecurity through AI and ML: Strategies, Challenges, and Future Directions," *Journal of Information Security* , vol. 15, no. 3, p. 320, Jan. 2024, doi: 10.4236/jis.2024.153019.

[32] S. Rao and D. Chatterjee, "Artificial Intelligence Policy:Need Aggressive Development with Prudent Regulation," *Current Science* , vol. 115, no. 6, p. 1015, Sep. 2018, doi: 10.18520/cs/v115/i6/1015-1016.

[33] M. Prakash, "The impact of artificial intelligence on the strategic power dynamics between the USA and India: A comparative analysis of technological advancements and geopolitical influence," *International Journal of Political Science and Governance* , vol. 6, no. 2, p. 220, Jul. 2024, doi: 10.33545/26646021.2024.v6.i2c.388.

[34] B. Guembe, S. Misra, A. Azeta, and I. López-Baldominos, "Bibliometric analysis of artificial intelligence cyberattack detection models," *Artificial Intelligence Review* , vol. 58, no. 6, Mar. 2025, doi: 10.1007/s10462-025-11167-0.

[35] N. Mohamed, "Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms," *Knowledge and Information Systems* , Apr. 2025, doi: 10.1007/s10115-025-02429-y.

[36] A. Shahana *et al.* , "AI-Driven Cybersecurity: Balancing Advancements and Safeguards," *Journal of Computer Science and Technology Studies* , vol. 6, no. 2, p. 76, May 2024, doi: 10.32996/jcsts.2024.6.2.9.

[37] S. Thapaliya and A. Bokani, "Leveraging artificial intelligence for enhanced cybersecurity: insights and innovations," *SADGAMAYA.* , vol. 1, no. 1, p. 46, Jun. 2024, doi: 10.3126/sadgamaya.v1i1.66888.

[38] A. O. Adewusi, U. I. Okoli, T. Olorunsogo, E. M. Adaga, D. O. Daraojimba, and O. Chimezie, "Artificial intelligence in cybersecurity: Protecting national infrastructure: A USA review," *World Journal of Advanced Research and Reviews* , vol. 21, no. 1. GSC Online Press, p. 2263, Jan. 27, 2024. doi: 10.30574/wjarr.2024.21.1.0313.

[39] C. Gilbert and M. A. Gilbert, "The Impact of AI on Cybersecurity Defense Mechanisms: Future Trends and Challenges," *SSRN Electronic Journal* , Jan. 2025, doi: 10.2139/ssrn.5258783.

[40] J. Oloyede, "AI-Driven Cybersecurity Solutions: Enhancing Defense Mechanisms in the Digital Era," *SSRN Electronic Journal* , Jan. 2024, doi: 10.2139/ssrn.4976103.

[41] B. T. Familoni, "CYBERSECURITY CHALLENGES IN THE AGE OF AI: THEORETICAL APPROACHES AND PRACTICAL SOLUTIONS," *Computer Science & IT Research Journal* , vol. 5, no. 3, p. 703, Mar. 2024, doi: 10.51594/csitrj.v5i3.930.

[42] A. T. Oyewole, C. C. Okoye, O. C. Ofodile, and C. E. Ugochukwu, "Cybersecurity risks in online banking: A detailed review and preventive strategies applicatio," *World Journal of Advanced Research and Reviews* , vol. 21, no. 3. GSC Online Press, p. 625, Mar. 11, 2024. doi: 10.30574/wjarr.2024.21.3.0707.

[43] C. Nobles and I. McAndrew, "The Intersectionality of Offensive Cybersecurity and Human Factors: A Position Paper," *Scientific Bulletin* , vol. 28, no. 2, p. 215, Dec. 2023, doi: 10.2478/bsaft-2023-0022.

[44] M. S. Urmila, "AI-Powered Cybersecurity for Critical Infrastructure: A Comprehensive Survey," *International Journal for Research in Applied Science and Engineering Technology* , vol. 13, no. 7, p. 2678, Jul. 2025, doi: 10.22214/ijraset.2025.73439.

[45] A. Banyal, "Synergies and Challenges: Exploring the Intersection of Artificial Intelligence and Cybersecurity," *International Journal for Research in Applied Science and Engineering Technology* , vol. 12, no. 4, p. 1091, Apr. 2024, doi: 10.22214/ijraset.2024.60021.

[46] K. N. Iyer, "Poisoning AI Models: New Frontiers in Data Manipulation Attacks," *International Journal of Innovative Research in Computer and Communication Engineering* , vol. 11, no. 11, Nov. 2023, doi: 10.15680/ijircce.2023.1111065.

[47] A. Kovačević, S. D. Radenković, and D. Nikolić, "Artificial intelligence and cybersecurity in banking sector: opportunities and risks," *arXiv (Cornell University)* , Nov. 2024, doi: 10.48550/arxiv.2412.04495.

[48] H. I. Kure, P. Sarkar, A. B. Ndanusa, and A. O. Nwajana, "Detecting and Preventing Data Poisoning Attacks on AI Models," *arXiv (Cornell University)* , Mar. 2025, doi: 10.48550/arxiv.2503.09302.

[49] M. Roshanaei, M. R. Khan, and N. N. Sylvester, "Navigating AI Cybersecurity: Evolving Landscape and Challenges," *Journal of Intelligent Learning Systems and Applications* , vol. 16, no. 3, p. 155, Jan. 2024, doi: 10.4236/jilsa.2024.163010.

[50] R. Huang, X. Zheng, Y. Shang, and X. Xue, "On challenges of AI to cognitive security and safety," *Security and Safety* , vol. 2, p. 2023012, Jan. 2023, doi: 10.1051/sands/2023012.

[51] B. Zweers, D. Dey, and D. Bhaumik, "The AI-Fraud Diamond: A Novel Lens for Auditing Algorithmic Deception," *arXiv (Cornell University)* , Aug. 2025, doi: 10.48550/arxiv.2508.13984.

[52] Y. Yigit *et al.* , "Critical Infrastructure Protection: Generative AI, Challenges, and Opportunities," *arXiv (Cornell University)* , May 2024, doi: 10.48550/arxiv.2405.04874.

[53] I. E. Kezron, "Securing the AI supply chain: Mitigating vulnerabilities in AI model development and deployment," *World Journal of Advanced Research and Reviews* , vol. 22, no. 2, p. 2336, May 2024, doi: 10.30574/wjarr.2024.22.2.1394.

[54] M. Raparthi, S. B. Dodda, and S. Maruthi, "Examining the use of Artificial Intelligence to Enhance Security Measures in Computer Hardware, including the Detection of Hardware-based Vulnerabilities and Attacks.," Jan. 2020, doi: 10.52783/eel.v10i1.991.

[55] B. Kereopa-Yorke, "Quantifying AI Vulnerabilities: A Synthesis of Complexity, Dynamical Systems, and Game Theory," *arXiv (Cornell University)* , Apr. 2024, doi: 10.48550/arxiv.2404.10782.

[56] A. Szabo and U. Hadad, "Crypto Miner Attack: GPU Remote Code Execution Attacks," *arXiv (Cornell University)* , Feb. 2025, doi: 10.48550/arxiv.2502.10439.

[57] A. Tsamados, L. Floridi, and M. Taddeo, "The Cybersecurity Crisis of Artificial Intelligence: Unrestrained Adoption and Natural Language-Based Attacks," *SSRN Electronic Journal* , Jan. 2023, doi: 10.2139/ssrn.4578165.

[58] M. H. Y. Binhammad, S. Alqaydi, A. Othman, and L. H. Abuljadayel, "The Role of AI in Cyber Security: Safeguarding Digital Identity," *Journal of Information Security* , vol. 15, no. 2, p. 245, Jan. 2024, doi: 10.4236/jis.2024.152015.

[59] G. Malik, R. Brahmbhatt, and P. Prashasti, "AI-Driven Security and Inventory Optimization: Automating Vulnerability Management and

Demand Forecasting in CI/CD-Powered Retail Systems," *International Journal of Computational and Experimental Science and Engineering* , vol. 11, no. 3, Sep. 2025, doi: 10.22399/ijcesen.3855.

[60] L. Alevizos and M. Dekker, "Towards an AI-Enhanced Cyber Threat Intelligence Processing Pipeline," *arXiv (Cornell University)* , Mar. 2024, doi: 10.3390/electronics13112021.

[61] H. Kheddar, "Transformers and Large Language Models for Efficient Intrusion Detection Systems: A Comprehensive Survey," *arXiv (Cornell University)* , Aug. 2024, doi: 10.48550/arxiv.2408.07583.

[62] V. Kulothungan, "Securing the AI Frontier: Urgent Ethical and Regulatory Imperatives for AI-Driven Cybersecurity," *arXiv (Cornell University)* , Jan. 2025, doi: 10.48550/arxiv.2501.10467.

[63] A. Kovačević, S. D. Radenković, and D. Nikolić, "ARTIFICIAL INTELLIGENCE AND CYBERSECURITY IN BANKING SECTOR: OPPORTUNITIES AND RISKS," *Zbornik radova - Geografski fakultet Univerziteta u Beogradu* , p. 425, Jan. 2025, doi: 10.46793/ebm24.425k.

[64] M. Shah, "A Social Outcomes and Priorities centered (SOP) Framework for AI policy," *arXiv (Cornell University)* , Nov. 2024, doi: 10.48550/arxiv.2411.08241.

[65] S. Banerjee, P. Sahu, M. Luo, A. Vahldiek-Oberwagner, N. J. Yadwadkar, and M. Tiwari, "SoK: A Systems Perspective on Compound AI Threats and Countermeasures," *arXiv (Cornell University)* , Nov. 2024, doi: 10.48550/arxiv.2411.13459.

[66] A. M. Elmisery, M. Sertovic, A. Zayin, and P. F. Watson, "Cyber Threats in Financial Transactions -- Addressing the Dual Challenge of AI and Quantum Computing," *arXiv (Cornell University)* , Mar. 2025, doi: 10.48550/arxiv.2503.15678.

[67] J. Zhang and D. Tenney, "The Evolution of Integrated Advance Persistent Threat and Its Defense Solutions: A Literature Review," *Open Journal of Business and Management* , vol. 12, no. 1. Scientific Research Publishing, p. 293, Jan. 01, 2024. doi: 10.4236/ojbm.2024.121021.

[68] M. Smith and J. Ingram, "Surveying the Operational Cybersecurity and Supply Chain Threat Landscape when Developing and Deploying AI Systems," *arXiv (Cornell University)* , Aug. 2025, doi: 10.48550/arxiv.2508.20307.

[69] H. Hayagreevan and S. Khamaru, "Security of and by Generative AI platforms," *arXiv (Cornell University)* , Oct. 2024, doi: 10.48550/arxiv.2410.13899.

[70] M. Grey and C.-R. Segerie, "The AI Risk Spectrum: From Dangerous Capabilities to Existential Threats," *arXiv (Cornell University)* , Aug. 2025, doi: 10.48550/arxiv.2508.13700.

[71] S. Abdali, R. Anarfi, C. Barberan, and J. He, "Securing Large Language Models: Threats, Vulnerabilities and Responsible Practices," *arXiv (Cornell University)* , Mar. 2024, doi: 10.48550/arxiv.2403.12503.

[72] D. Humphreys, A. Koay, D. Desmond, and E. Mealy, "AI hype as a cyber security risk: the moral responsibility of implementing generative AI in business," *AI and Ethics* , vol. 4, no. 3, p. 791, Feb. 2024, doi: 10.1007/s43681-024-00443-4.

[73] S. Girhepuje, A. Verma, and G. Raina, "A Survey on Offensive AI Within Cybersecurity," *arXiv (Cornell University)* , Sep. 2024, doi: 10.48550/arxiv.2410.03566.

[74] F. Jimmy, "Emerging Threats: The Latest Cybersecurity Risks and the Role of Artificial Intelligence in Enhancing Cybersecurity Defenses," *International Journal of Scientific Research and Management (IJSRM)* , vol. 9, no. 2, p. 564, Feb. 2021, doi: 10.18535/ijsrm/v9i2.ec01.

[75] O. S. Ajibola, O. Dopamu, and O. Olurin, "Challenges and ethical implications of using AI in cybersecurity," *International Journal of Science and Research Archive* , vol. 14, no. 2, p. 294, Feb. 2025, doi: 10.30574/ijsra.2025.14.2.0276.

[76] V. Singh and D. R. Gautam, "Cyber Crime, Security and Regulation in India," 2022, p. 147. doi: 10.55662/book.2022ccrs.005.

[77] R. Bharati, "Navigating the Legal Landscape of Artificial Intelligence: Emerging Challenges and Regulatory Framework in India," *SSRN Electronic Journal* , Jan. 2024, doi: 10.2139/ssrn.4898536.

[78] D. Cyman, E. Gromova, and E. Juchnevicius, "Regulation of Artificial Intelligence in BRICS and the European Union," *BRICS Law Journal*

, vol. 8, no. 1, p. 86, Apr. 2021, doi: 10.21684/2412-2343-2021-8-1-86-115.

[79] A. Daly *et al.* , "Artificial Intelligence, Governance and Ethics: Global Perspectives," *SSRN Electronic Journal* , Jan. 2019, doi: 10.2139/ssrn.3414805.

[80] N. Bhalla, L. Brooks, and T. Leach, "Ensuring a 'Responsible' AI future in India: RRI as an approach for identifying the ethical challenges from an Indian perspective," *AI and Ethics* , vol. 4, no. 4, p. 1409, Dec. 2023, doi: 10.1007/s43681-023-00370-w.

[81] D. I. Mikhailov, "Optimizing National Security Strategies through LLM-Driven Artificial Intelligence Integration," *arXiv (Cornell University)* , May 2023, doi: 10.48550/arxiv.2305.13927.

[82] S. Oh and M. R. Sanfilippo, "University Governance for Responsible AI," *Proceedings of the ALISE Annual Conference* , Oct. 2024, doi: 10.21900/j.alise.2024.1706.

[83] A. Ghosh, A. Saini, and H. Barad, "Artificial intelligence in governance: recent trends, risks, challenges, innovative frameworks and future directions," *AI & Society* , vol. 40, no. 7, p. 5685, Mar. 2025, doi: 10.1007/s00146-025-02312-y.

[84] I. T. Hjaltalin and H. T. Sigurdarson, "The strategic use of AI in the public sector: A public values analysis of national AI strategies," *Government Information Quarterly* , vol. 41, no. 1, p. 101914, Feb. 2024, doi: 10.1016/j.giq.2024.101914.

[85] B.-C. Ubaldi *et al.* , "Governing with Artificial Intelligence," Jun. 2024. doi: 10.1787/26324bc2-en.

[86] A. Aladiyan, "Digital Safeguards: Unravelling the Complex Interplay Between Emerging Threats and Proactive Cyber Defence Strategies," *Journal of Internet Services and Information Security* , vol. 15, no. 1, p. 348, Feb. 2025, doi: 10.58346/jisis.2025.i1.022.

[87] G. Kim and K. Park, "Effect of AI," *Tehnički glasnik* , vol. 18, no. 1, p. 29, Jan. 2024, doi: 10.31803/tg-20230218142012.

[88] M. Stoltz, "Artificial Intelligence in Cybersecurity: Building Resilient Cyber Diplomacy Frameworks," 2024, doi: 10.48550/ARXIV.2411.13585.

[89] A. Puczko, "An Ambiguous Relationship between Public Administration and AI," in *IntechOpen eBooks* , IntechOpen, 2024. doi: 10.5772/intechopen.115281.

[90] A. F. Vatamanu and M. Tofan, "Integrating Artificial Intelligence into Public Administration: Challenges and Vulnerabilities," *Administrative Sciences* , vol. 15, no. 4, p. 149, Apr. 2025, doi: 10.3390/admsci15040149.

[91] F. Pana-Micu, "ARTIFICIAL INTELLIGENCE IN THE PUBLIC SECTOR-CHALLENGES, OPPORTUNITIES AND BEST PRACTICES," *Journal of Public Administration Finance and Law* , vol. 32, p. 393, Jan. 2024, doi: 10.47743/jopafl-2024-32-29.

[92] M. Pietri, M. Mamei, and M. Colajanni, "Telecom spam and scams in the 5G and artificial intelligence era: analyzing economic implications, technical challenges and global regulatory efforts," *International Journal of Information Security* , vol. 24, no. 3, May 2025, doi: 10.1007/s10207-025-01062-8.

[93] A. Wang, S. Kapoor, S. Barocas, and A. Narayanan, "Against Predictive Optimization: On the Legitimacy of Decision-making Algorithms That Optimize Predictive Accuracy," *ACM Journal on Responsible Computing* , vol. 1, no. 1, p. 1, Dec. 2023, doi: 10.1145/3636509.

[94] A. Oesterling, U. Bhalla, S. Venkatasubramanian, and H. Lakkaraju, "Operationalizing the Blueprint for an AI Bill of Rights: Recommendations for Practitioners, Researchers, and Policy Makers," *arXiv (Cornell University)* , Jul. 2024, doi: 10.48550/arxiv.2407.08689.

[95] M. J. Ahn and Y. Chen, "Artificial Intelligence in Government:," p. 243, Jun. 2020, doi: 10.1145/3396956.3398260.

[96] S. Bhale, "Deepfake Laws in India: The Need for Legal Regulation in the AI Era," *SSRN Electronic Journal* , Jan. 2025, doi: 10.2139/ssrn.5153296.

[97] T. Urtmelidze, "The role of intellectual property in addressing state security challenges," *DergiPark (Istanbul University)* , Jan. 2025, Accessed: Oct. 2025. [Online]. Available: https://dergipark.org.tr/en/pub/usbed/issue/88154/1623886

[98] N. Deo and P. A. Singh, "Cybersecurity and Sustainable Development," 2022, p. 188. doi: 10.55662/book.2022ccrs.009.

[99] F. Schiliro, "From Crime to Hypercrime: Evolving Threats and Law Enforcement's New Mandate in the AI Age," *arXiv (Cornell University)* , Nov. 2024, doi: 10.48550/arxiv.2411.10995.

[100] A. V. Priya, "CRIMINAL ACCOUNTABILITY FOR AI: MENS REA, ACTUS REUS, AND THE CHALLENGES OF AUTONOMOUS SYSTEMS," *LawFoyer International Journal of Doctrinal Legal Research.* , vol. 3, no. 1, p. 273, Apr. 2025, doi: 10.70183/lijdlr.2024.v03.13.

[101] D. U. S. de la Osa and N. Remolina, "Artificial intelligence at the bench: Legal and ethical challenges of informing—or misinforming—judicial decision-making through generative AI," *Data & Policy* , vol. 6, Jan. 2024, doi: 10.1017/dap.2024.53.

[102] S. Mandal, "Deep Fake Technology and Identity Theft: An Emerging Challenge for Cyber Laws in India," Jan. 2025, doi: 10.2139/ssrn.5161545.

[103] K. Garimella and S. Chauchard, "Is AI misinformation influencing elections in India?," *Nature* , vol. 630, no. 8015, p. 32, Jun. 2024, doi: 10.1038/d41586-024-01588-2.

[104] A. D. Samuel-Okon, O. O. Olateju, S. U. Okon, O. O. Olaniyi, and U. T. I. Igwenagu, "Formulating Global Policies and Strategies for Combating Criminal Use and Abuse of Artificial Intelligence," *Archives of Current Research International* , vol. 24, no. 5, p. 612, Jun. 2024, doi: 10.9734/acri/2024/v24i5735.

[105] D. W. Linna *et al.* , "Deepfakes in Court: How Judges Can Proactively Manage Alleged AI-Generated Material in National Security Cases," Jan. 2024, doi: 10.2139/ssrn.4943841.

[106] A. Mukherjee and S. P. Chakrabarty, "Can AI Help Indian Judiciary to Reduce Its Burden of Cases? Exploring the Potential of AI in Judicial Decision-Making Process," in *Lecture notes in networks and systems* , Springer International Publishing, 2025, p. 467. doi: 10.1007/978-981-97-8457-8_29.

[107] M. Casu, L. Guarnera, P. Caponnetto, and S. Battiato, "GenAI Mirage: The Impostor Bias and the Deepfake Detection Challenge in the Era of Artificial Illusions," *arXiv (Cornell University)* , Dec. 2023, doi: 10.48550/arxiv.2312.16220.

[108] M. K. Sharma, "India's Courts and Artificial Intelligence: A Future Outlook," *Lexonomica* , vol. 15, no. 1, Jan. 2023, doi: 10.18690/lexonomica.15.1.99-120.2023.

[109] M. Sokolova, "The Role and Position of AI Evidence in Civil Litigation," *Medicine Law & Society* , vol. 16, no. 1, Jan. 2023, doi: 10.18690/mls.16.1.169-190.2023.

[110] Y. Chen, "When Algorithms Testify: Addressing the Explainability Gap of AI Evidence in Criminal Cases," *Studies in Law and Justice* , vol. 4, no. 3, p. 1, May 2025, doi: 10.56397/slj.2025.06.01.

[111] DR. N. SHARMA, "ARTIFICIAL INTELLIGENCE: LEGAL IMPLICATIONS AND CHALLENGES," *Knowledgeable Research A Multidisciplinary Journal* , vol. 2, no. 11, p. 13, Jun. 2024, doi: 10.57067/220k4298.

[112] N. Madaoui, "The Impact of Artificial Intelligence on Legal Systems: Challenges and Opportunities," *Problems of Legality* , vol. 1, no. 164, p. 285, May 2024, doi: 10.21564/2414-990x.164.289266.

[113] A. Balahur *et al.* , "Data quality requirements for inclusive, non-biased and trustworthy AI," *HAL (Le Centre pour la Communication Scientifique Directe)* , Dec. 2022, doi: 10.2760/365479.

[114] K. Javed and J. Li, "Bias in adjudication: Investigating the impact of artificial intelligence, media, financial and legal institutions in pursuit of social justice," *PLoS ONE* , vol. 20, no. 1, Jan. 2025, doi: 10.1371/journal.pone.0315270.

[115] A. Fine, E. R. Berthelot, and S. Marsh, "Public Perceptions of Judges' Use of AI Tools in Courtroom Decision-Making: An Examination of Legitimacy, Fairness, Trust, and Procedural Justice," *Behavioral Sciences* , vol. 15, no. 4, p. 476, Apr. 2025, doi: 10.3390/bs15040476.

[116] C. Madhumitha, "Transition from Human to AI-Judge - The Future Automated Adjudication Mechanism in India," *International Journal for Research in Applied Science and Engineering Technology* , vol. 12, no. 11, p. 540, Nov. 2024, doi: 10.22214/ijraset.2024.65041.

[117] A. R. Vargas-Murillo, I. N. M. de la A. Pari-Bedoya, A. M. Turriate-Guzmán, C. A. Delgado-Chávez, and F. Sanchez-Paucar,

"Transforming Justice: Implications of Artificial Intelligence in Legal Systems," *Academic Journal of Interdisciplinary Studies* , vol. 13, no. 2, p. 433, Mar. 2024, doi: 10.36941/ajis-2024-0059.

[118] Y. Abuzir, "Artificial Intelligence in Legal Practice: Applications, Challenges, and Future Prospects," *DergiPark (Istanbul University)* , Jan. 2025, doi: 10.46238/jobda.1629307.

[119] K. Verma, "Analyzing the Endeavours of the Supreme Court of India to Transcribe and Translate Court Arguments in Light of the Proposed EU AI Act," *arXiv (Cornell University)* , Sep. 2023, doi: 10.48550/arxiv.2309.10088.

[120] A. Vats, "Building the case for restricted use of predictive policing tools in India," *The International Review of Information Ethics* , vol. 32, no. 1, Nov. 2022, doi: 10.29173/irie487.

[121] R. Faqir, "THE EXCLUSIONARY RULE OF AI-ENHANCED DIGITAL EVIDENCE IN THE UNITED STATES AND UAE: A COMPARATIVE ANALYSIS," *Journal of Southwest Jiaotong University* , vol. 59, no. 1, Jan. 2024, doi: 10.35741/issn.0258-2724.59.1.7.

[122] P. Gawali and R. Sony, "The Role of Artificial Intelligence in Improving Criminal Justice System: Indian Perspective," *Legal Issues in the Digital Age* , vol. 3, no. 3, p. 78, Dec. 2020, doi: 10.17323/2713-2749.2020.3.78.96.

[123] T. Odelberg, "Artificial Intelligence Handbook for Local Government," *Deep Blue (University of Michigan)* , Sep. 2024, doi: 10.7302/24632.

[124] I. Calzada, G. Németh, and M. S. Al-Radhi, "Trustworthy AI for Whom? GenAI Detection Techniques of Trust Through Decentralized Web3 Ecosystems," *Big Data and Cognitive Computing* , vol. 9, no. 3, p. 62, Mar. 2025, doi: 10.3390/bdcc9030062.

[125] I. Cheong, A. Caliskan, and T. Kohno, "Safeguarding human values: rethinking US law for generative AI's societal impacts," *AI and Ethics* , May 2024, doi: 10.1007/s43681-024-00451-4.

[126] M. Roshan, "Generative AI in Fintech: Advancing Risk Assessment and Fraud Detection in Digital Payment Technologies," *International Journal for Research in Applied Science and Engineering Technology* , vol. 12, no. 8, p. 1318, Aug. 2024, doi: 10.22214/ijraset.2024.64110.

[127] N. Li, "Ethical Considerations in Artificial Intelligence: A Comprehensive Disccusion from the Perspective of Computer Vision," *SHS Web of Conferences* , vol. 179, p. 4024, Jan. 2023, doi: 10.1051/shsconf/202317904024.

[128] E. Kurshan, D. Mehta, B. Bruss, and T. Balch, "AI versus AI in Financial Crimes and Detection: GenAI Crime Waves to Co-Evolutionary AI," *arXiv (Cornell University)* , Sep. 2024, doi: 10.48550/arxiv.2410.09066.

[129] N. Yalçın and B. Lale, "Types of cyber-attacks with using voice," *DergiPark (Istanbul University)* , Dec. 2024, Accessed: Oct. 2025. [Online]. Available: https://dergipark.org.tr/en/pub/jsr-a/issue/93120/1600934

[130] E. Ferrara, "GenAI against humanity: nefarious applications of generative artificial intelligence and large language models," *Journal of Computational Social Science* , vol. 7, no. 1, p. 549, Feb. 2024, doi: 10.1007/s42001-024-00250-1.

[131] E. Ferrara, "GenAI Against Humanity: Nefarious Applications of Generative Artificial Intelligence and Large Language Models," *SSRN Electronic Journal* , Jan. 2023, doi: 10.2139/ssrn.4614223.

[132] Y. Bengio *et al.* , "International AI Safety Report," *arXiv (Cornell University)* , Jan. 2025, doi: 10.48550/arxiv.2501.17805.

[133] G. Mittal, A. Jakobsson, K. O. Marshall, C. Hegde, and N. Memon, "AI-assisted Tagging of Deepfake Audio Calls using Challenge-Response," *arXiv (Cornell University)* , Feb. 2024, doi: 10.48550/arxiv.2402.18085.

[134] S. M. Jayakannan, "Securing Voice-Based Financial Authentication in the Era of AI Voice Cloning: Challenges, Vulnerabilities, and Counter-Measures," *Journal of Computer Science and Technology Studies* , vol. 7, no. 4, p. 515, May 2025, doi: 10.32996/jcsts.2025.7.4.60.

[135] K. Huang and B. Hu, "Hybrid Audio Detection Using Fine-Tuned Audio Spectrogram Transformers: A Dataset-Driven Evaluation of Mixed AI-Human Speech," *arXiv (Cornell University)* , May 2025, doi: 10.48550/arxiv.2505.15136.

[136] S. Ahmadi, "Open AI and its Impact on Fraud Detection in Financial Industry," *Journal of Knowledge Learning and Science Technology ISSN 2959-6386 (online)* , vol. 2, no. 3, p. 263, Jan. 2024, doi: 10.60087/jklst.vol2.n3.p281.

[137] W. Guo, Y. Potter, T. Shi, Z. Wang, A. Zhang, and D. Song, "Frontier AI's Impact on the Cybersecurity Landscape," *arXiv (Cornell University)* , Apr. 2025, doi: 10.48550/arxiv.2504.05408.

[138] M. Schmitt and I. Fléchais, "Digital Deception: Generative Artificial Intelligence in Social Engineering and Phishing," *SSRN Electronic Journal* , Jan. 2023, doi: 10.2139/ssrn.4602790.

[139] M. Schmitt and I. Fléchais, "Digital deception: generative artificial intelligence in social engineering and phishing," *Artificial Intelligence Review* , vol. 57, no. 12, Oct. 2024, doi: 10.1007/s10462-024-10973-2.

[140] M. Verma, "AI-Driven Cyber Attacks and 5G Networks: The New Age of Digital Threats," *International Journal for Research in Applied Science and Engineering Technology* , vol. 13, no. 2, p. 1469, Feb. 2025, doi: 10.22214/ijraset.2025.67117.

[141] M. Rodriguez, R. A. Popa, F. Flynn, L. Liang, A. Dafoe, and A. Wang, "A Framework for Evaluating Emerging Cyberattack Capabilities of AI," 2025, doi: 10.48550/ARXIV.2503.11917.

[142] Y. Ge and Q. Zhu, "The Game-Theoretic Symbiosis of Trust and AI in Networked Systems," *arXiv (Cornell University)* , Nov. 2024, doi: 10.48550/arxiv.2411.12859.

[143] J. Park, H. Chung, and J. F. DeFranco, "Multilayered Diagnostics for Smart Cities," *Computer* , vol. 55, no. 2, p. 14, Feb. 2022, doi: 10.1109/mc.2021.3070325.

[144] S. Vempati, "Securing Smart Cities: A Cybersecurity Perspective on Integrating IoT, AI, and Machine Learning for Digital Twin Creation," *Deleted Journal* , vol. 20, no. 3, p. 1420, May 2024, doi: 10.52783/jes.3548.

[145] E. Khalifa, "The Impact of Smart City Model on National Security," *Central European Journal of International and Security Studies* , vol. 14, no. 1, p. 52, Mar. 2020, doi: 10.51870/cejiss.a140103.

[146] J. S. Oliha, P. W. Biu, and O. Chimezie, "SECURING THE SMART CITY: A REVIEW OF CYBERSECURITY CHALLENGES AND STRATEGIES," *Engineering Science & Technology Journal* , vol. 5, no. 2. Fair East Publishers, p. 496, Feb. 25, 2024. doi: 10.51594/estj.v5i2.827.

[147] E. O. Udeh, P. Amajuoyi, K. B. Adeusi, and A. O. Scott, "The integration of artificial intelligence in cybersecurity measures for sustainable finance platforms: An analysis," *Computer Science & IT Research Journal* , vol. 5, no. 6, p. 1221, Jun. 2024, doi: 10.51594/csitrj.v5i6.1195.

[148] U. M. Adanma and E. O. Ogunbiyi, "Artificial intelligence in environmental conservation: evaluating cyber risks and opportunities for sustainable practices," *Computer Science & IT Research Journal* , vol. 5, no. 5, p. 1178, May 2024, doi: 10.51594/csitrj.v5i5.1156.

[149] S. K. Srivastava, "AI in Justice Delivery in India: Current Status, Identification of Potential Applications and Way Forward," *Informatica* , vol. 47, no. 5, May 2023, doi: 10.31449/inf.v47i5.4361.

[150] T. Saheb and T. Saheb, "Mapping Ethical Artificial Intelligence Policy Landscape: A Mixed Method Analysis," *Science and Engineering Ethics* , vol. 30, no. 2, Mar. 2024, doi: 10.1007/s11948-024-00472-6.

[151] A. Sachoulidou, "Going beyond the 'common suspects': to be presumed innocent in the era of algorithms, big data and artificial intelligence," *Artificial Intelligence and Law* , Feb. 2023, doi: 10.1007/s10506-023-09347-w.

[152] C. Moore *et al.* , "Concerning the Responsible Use of AI in the US Criminal Justice System," *arXiv (Cornell University)* , May 2025, doi: 10.48550/arxiv.2506.00212.

[153] S. Ghosh, D. Verma, B. Ganesan, P. Bindal, V. Kumar, and V. Bhatnagar, "Human Centered AI for Indian Legal Text Analytics," *arXiv (Cornell University)* , Mar. 2024, doi: 10.48550/arxiv.2403.10944.

[154] S. Girhepuje *et al.* , "Are Models Trained on Indian Legal Data Fair?," *arXiv (Cornell University)* , Jan. 2023, doi: 10.48550/arxiv.2303.07247.

[155] O. Pakuhinezhad and A. Atrian, "Navigating the Ethical, Societal, and Technological Challenges of AI Judges: Toward Responsible Legal Systems," *SSRN Electronic Journal* , Jan. 2024, doi: 10.2139/ssrn.4816558.

[156] A. Zafar, "Balancing the scale: navigating ethical and practical challenges of artificial intelligence (AI) integration in legal practices,"

Discover Artificial Intelligence , vol. 4, no. 1, Apr. 2024, doi: 10.1007/s44163-024-00121-8.

[157] D. Byrd, "A+AI: Threats to Society, Remedies, and Governance," *arXiv (Cornell University)* , Sep. 2024, doi: 10.48550/arxiv.2409.02219.

[158] P. Andrews *et al.* , "A Trust Framework for Government Use of Artificial Intelligence and Automated Decision Making," *arXiv (Cornell University)* , Aug. 2022, doi: 10.48550/arxiv.2208.10087.

[159] M. K. S. Warraich, H. Usman, S. Zakir, and M. Mehboob, "Ethical Governance of artificial intelligence Hallucinations in legal practice," vol. 4, no. 2, p. 603, May 2025, doi: 10.71085/sss.04.02.297.

[160] E. Schneiders *et al.* , "Objection Overruled! Lay People can Distinguish Large Language Models from Lawyers, but still Favour Advice from an LLM," *arXiv (Cornell University)* , Sep. 2024, doi: 10.48550/arxiv.2409.07871.

[161] N. K. Corrêa and J. M. Mönig, "Catalog of General Ethical Requirements for AI Certification," *arXiv (Cornell University)* , Aug. 2024, doi: 10.48550/arxiv.2408.12289.

[162] N. R. Barman *et al.* , "The Brittleness of AI-Generated Image Watermarking Techniques: Examining Their Robustness Against Visual Paraphrasing Attacks," *arXiv (Cornell University)* , Aug. 2024, doi: 10.48550/arxiv.2408.10446.

[163] O. Ritchie, M. Anderljung, and T. Rachman, "From Turing to Tomorrow: The UK's Approach to AI Regulation," *arXiv (Cornell University)* , Jul. 2025, Accessed: Oct. 2025. [Online]. Available: http://arxiv.org/abs/2507.03050

[164] A. Reuel *et al.* , "Open Problems in Technical AI Governance," *arXiv (Cornell University)* , Jul. 2024, doi: 10.48550/arxiv.2407.14981.

[165] X. Zhao *et al.* , "SoK: Watermarking for AI-Generated Content," *arXiv (Cornell University)* , Nov. 2024, doi: 10.48550/arxiv.2411.18479.

[166] A. Nemecek, Y. Jiang, and E. Ayday, "Watermarking Without Standards Is Not AI Governance," *arXiv (Cornell University)* , May 2025, doi: 10.48550/arxiv.2505.23814.

[167] B. Rijsbosch, G. van Dijck, and K. Kollnig, "Adoption of Watermarking Measures for AI-Generated Content and Implications under the EU AI Act," 2025, doi: 10.48550/ARXIV.2503.18156.

[168] M. Christodorescu *et al.* , "Securing the Future of GenAI: Policy and Technology," *arXiv (Cornell University)* , May 2024, doi: 10.48550/arxiv.2407.12999.