AI-DRIVEN CTI FOR BUSINESS: EMERGING THREATS, ATTACK STRATEGIES, AND DEFENSIVE MEASURES

Mohammed Kashif [1], Mohammed Aasimuddin [2], Mubashir Ali Ahmed [3], Laxmi Bhavani Cheekatimalla [4], Eraj Farheen Ansari [5], Ahwan Mishra [6]

[1] Department of IT, Jamia Millia Islamia University, New Delhi, India,
[2] Department of IT, Campbellsville University, KY, USA
[3] Department of Computer Science, University of the People, CA, USA
[4] Department of Mathematics, Kakatiya University, India
[5] Department of Business Administration, East-West University, IL, USA
[6] Denmark High School, Alpharetta, GA, USA

ABSTRACT

Artificial Intelligence (AI) has transformed multiple sectors but also introduced new risks through its use in sophisticated cybercrimes. AI-driven cyberattacks leverage machine learning, deep learning, and natural language processing to enhance the scale, speed, and precision of threats like phishing, malware, identity theft, and social engineering. This paper examines AI's dual role—both as an enabler of complex cyberattacks and a tool for cybersecurity defence. Real-world examples include deepfake scams, heuristic malware, and AI-powered reconnaissance tools utilized by cybercriminals employing techniques such as Generative Adversarial Networks (GANs) and reinforcement learning to bypass conventional security systems. The dynamic nature of AI-enabled threats complicates detection and response, as these attacks adapt in real time. Conversely, AI is also applied defensively for predictive threat modelling, anomaly detection, behaviour-based intrusion prevention, and rapid incident response. The research underscores the necessity of robust AI governance, ethical AI use, and international collaboration to address these challenges. It concludes by proposing a strategic roadmap for stakeholders to develop AI-aware cybersecurity systems capable of evolving alongside the continuously changing threat landscape, aiming to harness AI's potential for protection while mitigating its exploitation for malicious purposes.

KEYWORDS

DUET, Cybercrime, Deepfake, Phishing, CTI, GANs, Social Engineering, Cybersecurity, AI Attacks, Defensive AI, AI Governance, Adversarial AI, Threat Detection, AI Ethics, AI Vulnerabilities

1. INRODUCTION

Artificial Intelligence (AI) has emerged as a game-changer in Cyber Threat Intelligence (CTI) techniques where large volumes of data is handled, predictions are made, and systems are automated [1]. AI has led to innovation and effectiveness in health and finance, communications and logistics, etc. With all its positive uses, AI has created an illusory, negative side—as applied to crime. The most concerning trend in this regard is the creation of AI-enabled cybercrime, where high-level algorithms are employed to multiply the quantity, speed, precision, and sophistication of cyberattacks [2].

Cybercrime used to rely heavily on human efforts and low-level automation. Now, with the emergence of advanced machine learning algorithms, neural networks, and natural language

processing (NLP), there has never been it as good as today for cybercriminals [3]. They enable hackers to craft highly advanced phishing emails, forge fake identities with deepfakes, scan vulnerabilities automatically, and conduct persistent, adaptive malware attacks with minimal human involvement. The result is an increasing cyber threat landscape that equally targets individuals, organizations, and governments.

The shift to AI-driven cyberattacks is more than a change of methodology; it is an inherent revolution within the nature of cyber-attacks [4]. AI can scan vast amounts of data to locate vulnerabilities, evade detection by masquerading as normal user traffic, and self-improve continuously based on system responses. These have created new forms of attack, ranging from AI-based disinformation to automated spear phishing and self-replicating malware with reinforcement learning capabilities [5]. They are more evasive, more malicious, and more targeted than traditional attacks and hence more challenging to detect and block.

While AI is therefore so promising as a weapon of cyber offense, it is similarly vital as a weapon of cyber defence [6]. Under ethical and responsible use, AI can be a priceless ally in threat detection, behaviour-based analytics, and incident response in real-time [7]. While the cat-and-mouse race between offense and defence in AI capabilities is speeding up, it underscores the imperative for strong regulation frameworks, global cooperation, and ethical standards [8].

This research paper seeks to explore comprehensively the field of AI-based cybercrime, both offensively and defensively encompassing AI. It will present case studies from real-life, most meaningful methodologies, and what the implications are for cybersecurity practitioners and policymakers [9]. The paper also addresses the ethical, legal, and regulatory implications that emerge due to adversarial AI, and presents strategic recommendations on how to minimize the threat. By shedding light on the potential and danger of AI in cybercrime, the research aims to create a stronger understanding of how society will prepare and react to this multifaceted and evolving threat.

2. LITERATURE REVIEW

The body of literature on cybercrime involving AI has expanded in the recent decades with growing concern over the dual-use potential of artificial intelligence. This section discusses prevailing literature considering the application of AI for cybercrime, illustrates the evolving nature of threats, and determines gaps in studies to be covered [10].

2.1. AI as a Facilitator of Advanced Cyberattacks

Among the earliest of such papers is Brundage et al. (2018), who discussed the misuse of AI and warned that AI would be used for grand cyberattacks, most notably in the form of phishing and data poisoning using machine learning [11]. They pointed out how ML models can be trained to mimic normal behaviour and hence the attacks would be effectively impossible to spot. Likewise, Buchanan et al. (2020) stated that AI provides attacks that are more adaptive, persistent, and scalable in comparison to traditional methods [12].

2.2. Deepfake Technology and Identity Fraud

The growing application of deep learning, particularly Generative Adversarial Networks (GANs), has attracted the attention of scholars like Chesney and Citron (2019) [13]. They investigated how deepfakes are seriously affecting individuals' security, propaganda, and deception, specifically in finance and politics. Kietzmann et al. (2020) also analysed the impact of deepfakes

on business and cybercrime, illustrating how AI-generated video and audio can impersonate real actors with great fidelity [14].

2.3. Social Engineering Phishing through Artificial Intelligence

Phishing and social engineering through NLP models were thoroughly investigated by Kumar and Singh (2021), who highlighted how phishing emails can be made extremely authentic with the help of GPT-based models [15]. They demonstrated, through their experimentation, that AI-powered phishing attacks are extremely effective since they provide contextual personalization. Williams et al. (2022) also researched AI-based chatbots utilized in social engineering to deceive users into sharing confidential information.

2.4. Defensive AI Applications in Cybersecurity

Conversely, a study by Sommer and Brown (2021) investigated the use of AI for defence, including anomaly detection and threat intelligence systems. They described AI-based Security Information and Event Management (SIEM) systems that can identify abnormal behaviour and risk elimination in real time. In addition, Sharmeen et al. [16] investigated reinforcement learning models that defend adaptively against zero-day attacks by learning from the attack pattern.

2.5. Impacts of Current Research

While existing literature reveals the offence and defence roles played by AI, there are gaps in regulatory policies, ethical abuse of AI, and enforcement across borders. Few works exist that discuss frameworks of international cooperation or threat taxonomies overall for collective action against AI-enabled cyber threats [17].

3. TYPES OF AI-DRIVEN CYBERCRIMES

The intersection of artificial intelligence and cybercrime methods has brought with it emerging threat vectors that are more evasive, advanced, and scalable compared to the past [18]. These are the key groups of AI-based cybercrime, using some AI technologies to perform or augment malicious activity.

3.1. Deepfake Scams and Identity Fraud

Deepfakes, created via Generative Adversarial Networks (GANs), are the most conspicuous AI-powered threats [19]. Such artificial media entities can convincingly mimic public officials, CEOs, or common individuals and are commonly utilized in financial scams, disinformation operations, or smear campaigns. For instance, deepfake voice scams have triggered unauthorized wire transfer by mimicking company executives.

3.2. AI-based Phishing and Spear Phishing

Phishing has been upgraded from bulk-email spam to very high-level attacks. With Natural Language Processing (NLP) and language generation algorithms such as GPT, now attackers can develop context-based messages that are similar to actual communication [20]. Such phishing attacks are almost impossible to differentiate from actual correspondence, and therefore credential harvesting and financial fraud success rates have improved.

3.3. Autonomous Vulnerability Scanning and Exploit Generation

AI tools can be programmed to scan systematically for system vulnerabilities, identify poor configurations, and even generate tailor-made exploits. Autonomous tools like these minimize time spent in manual reconnaissance and attack planning, which helps low-capability attackers perform advanced intrusions with less effort [21].

3.4. Social Engineering Bots

Artificially intelligent chatbots and voice assistants can respond to victims in real time and extract sensitive data by simulating natural user activity. These AI applications can manipulate users by bullying through context-specific phrases and emotional influence, thereby becoming useful in customer support scams, computer support scams, and identity theft [22].

3.5. Adaptive Malware and Polymorphic Threats

Malware has evolved, and others employ machine learning to dynamically alter their signatures and behaviours so they will not be detected. These polymorphic threats evolve based on the victim's environment, changing their patterns and modes of approach so that they evade detection by antivirus tools and intrusion detection systems [23].

Bar Diagram: Prevalence of AI-Driven Cybercrime Types (2025 Projection)

The following is a bar graph showing the estimated prevalence (in %) of each of the main forms of AI-powered cybercrime in 2025 based on collated forecasts from the cybersecurity world:

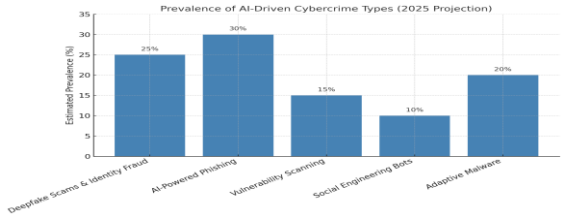

Fig 1: Prevalence of AI-Driven Cybercrime Types (2025 Projection)

4. LLM-ATTACK TECHNIQUES

AI-powered cyber-attacks employ different sophisticated techniques to infiltrate systems, hijack users, and bypass conventional defence mechanisms [24]. These techniques are borrowed from machine learning (ML), natural language processing (NLP), reinforcement learning (RL), and deep learning architectures to design highly efficient and adaptive attack models [25].

The primary techniques employed are explained below:

4.1. Pattern Recognition and Attack Targeting Through Machine Learning

Machine learning algorithms are employed to browse through massive amounts of information in order to discover possible targets and anticipate user actions [26]. Attackers utilize supervised and unsupervised learning to profile the victim, choose patterns from social networking sites or network activity logs, and develop successful spear phishing or credential stuffing attacks. By

recognizing repeating behaviours or system vulnerabilities, these applications make the attacks more accurate and streamlined [27].

4.2. Natural Language Processing (NLP) for Content Generation

NLP allows attackers to create indistinguishable text from human-written text employed in phishing emails, chatbot scams, or impersonation messages [28]. One can train models such as GPT-4 or BERT to replicate corporate tone, understand context, and compose extremely believable emails or messages that are hard to recognize as spoofed in comparison to actual communication [29]. These techniques make social engineering significantly more efficient and scalable.

4.3. Generative Adversarial Networks (GANs) for Deepfakes and Data Forgery

GANs are two neural networks — discriminator and generator — in a battle to create progressively realistic synthetic content. The process is commonly used to generate deepfake images, audio, or video to impersonate real people or authenticate official documents [30]. GANs are also utilized to contaminate training data sets or create biometric authentication inputs such as facial scans or prints.

4.4. Reinforcement Learning for Adaptive Intrusion

RL allows AI agents to learn policies by mere explorations of the environment. In cybercrime, RL is used to attempt automatic intrusion through trial and error [31]. RL agents, for example, can evade multi-factor authentication mechanisms, employ firewalls, or dynamically modify malware payloads based on the feedback of a system to learn from every try to become better.

4.5. AI in Malware Obfuscation and Evasion

AI methods assist malware to bypass detection using adversarial ML methods. Malware attackers construct adversarial instances which are innocent to a machine learning system but embedded with covert malicious payloads [32]. Malware also utilizes AI to modify its binary signature or behaviour upon detection, thereby transforming in real-time to evade defence mechanisms.

Diagram: AI-Driven Cyberattack Lifecycle

The following diagram illustrates how different AI techniques are involved in every stage of a cyberattack — from reconnaissance to post-exploitation:

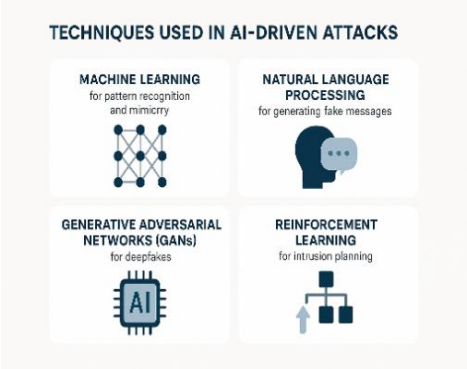

Fig 2: AI-Driven Cyberattack Lifecycle

5. CASE STUDIES AND CONTRIBUTIONS

The use of AI in cybercrime has evolved from threat theory to established fact. This chapter provides state-of-the-art real-world cases by industry, showing how attackers used artificial intelligence to mimic, take advantage of, and disrupt [33]. Each case demonstrates unique attack vectors, impact, and applicability for cybersecurity [34].

Contributions.

The technique used here is DUET—a novel framework called Distilled Uncertainty Ensemble for Threat detection. DUET integrates a supervised classification model with an unsupervised anomaly detection system, enhanced by calibrated uncertainty scoring, to effectively identify both known and emerging threats, including zero-day and polymorphic attacks. We developed a prototype and tested it on several benchmark datasets: UNSW-NB15, CIC-IDS2017, CSE-CIC-IDS2018 (focused on network intrusions), and SOREL-20M/EMBER (targeting polymorphic malware). DUET's performance was compared against existing methods such as Kitsune (autoencoder ensemble), LightGBM, Random Forest, MalConv (CNN-based), and BERT-style phishing detectors. Evaluation metrics included AUROC, PR-AUC, TPR at low FPR thresholds, and detection latency under realistic data splits. DUET consistently achieved higher true positive rates at low false positive rates (0.1%–1%) and demonstrated superior generalization to novel and distribution-shifted attacks compared to single-model approaches.

Proposed Method.

DUET has been extended to work with LLMs. It has a supervised classifier, unsupervised normality model and uncertainty scoring. The detail flow chart is described below.

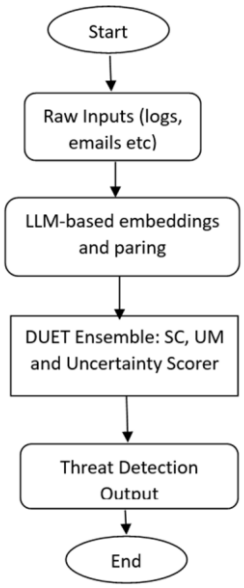

Fig 3. Flow chart of DUET + LLM process

5.1. Deepfake CEO Voice Scam (UK, 2019)

One of the earliest headline-making AI-powers cyberattacks was in 2019 when a UK power company's chief executive was duped into transferring a €220,000 payment by a call pretending to be from the parent company's German boss [35]. The voice, generated with deepfake technology, instructed the payment to be made to a counterfeit Hungarian supplier. An attacker used AI-based voice cloning to bypass usual identification checks, a new financial fraud modus operandi.

5.2. AI-Driven Phishing Attacks (Global, 2020–2023)

Between 2020 and 2023, some cybercrime gangs, e.g., FIN7 and APT33, had been using AI-based phishing generators to execute huge, multi-language campaigns [36]. The attacks were tailored via NLP and machine learning to take advantage of user behaviour, corporate jargon, and even regional cultural practices. The success rate of phishing is said to have been improved by over 40% compared to the traditional method [37].

5.3. Autonomous Vulnerability Scanning in SolarWinds Hack (USA, 2020)

Although not entirely AI-powered, experts reported that the SolarWinds attack modules showed ML uses optimized automation recon tools [38]. Evasiveness of the attackers over multiple layers and adaptive adjustment of their behaviour for the changing payload showed implementing adaptive algorithms [39]. The example shows the way stealth and persistence are facilitated in supply chain attacks through AI.

5.4. AI Chatbots Used in Tech Support Scams (India-USA, 2022)

Symantec exposed a gang of AI chatbots impersonating Microsoft and Apple technical support staff in 2022 [40]. The chatbots were able to communicate in real time with unsuspecting victims, identify "false" problems, and guide them to give remote access or buy useless software. NLP and intent detection formed the basis of their criminal conversation [41].

Table 1. Overview of AI-Driven Cybercrime Case Studies

Case	Year	AI Technique	Target Sector	Impact
Deepfake CEO Voice Scam (UK)	2019	Deepfake Voice Cloning	Finance	€220,000 lost via fraudulent wire
AI-Enhanced Phishing Campaigns (Global)	2020–23	NLP & ML Targeting	Multiple	40% rise in phishing success rate
SolarWinds Supply Chain Breach	2020	Automated ML Scanning	Government/Tech	Undetected intrusion for months
AI Chatbots in Tech Support Scams	2022	NLP-based Conversational AI	Consumer/IT Support	Widespread remote access fraud

6. CHALLENGES TO AI-BASED SOLUTIONS

AI brings game-changing cybersecurity technologies on the table but also comes with unprecedented challenges if applied in the wrong way. AI-powered cyberattacks are more scalable, responsive, and evasive than ever [42]. Detection and prevention require a new

paradigm both technology and strategy meant. The following section records the biggest challenges in different dimensions**.**

6.1. Avoidance of Traditional Security Controls

Threats produced by AI will be in the manner of emulating typical behaviour and it is difficult to detect using rule-based firewalls, antivirus programs, or signature-based IDS. Adaptive malware, for instance, will change its payload or encryption dynamically in order to avoid static analysis, whereas spam filters are bypassed by AI-generated phishing emails by mimicking writing style and context [43].

6.2. Real-Time Learning and Adaptation

Unlike fixed attacks, cyberattacks enabled by machine learning have the capability to learn. Through adversarial training or reinforcement learning, an offensive party can learn to optimize their strategy according to the defence mechanisms implemented. Real-time learning reduces the window of detection and enhances the hit rate of attacks considerably [44].

6.3. Volume and Velocity of Attacks

AI facilitates large-scale production and dissemination of malware, allowing attackers to scale and build their campaigns more than hitherto precedent [45]. For example, AI can create hundreds of thousands of spoofed accounts, phishing copies, or attack payloads in a matter of seconds, inundating security teams and increasing the likelihood of penetration.

6.4. In explainability of AI Detection Systems

While AI is also applied in defence fashion, black-box models such as deep neural networks are typically not explainable [46]. Security experts do not always know why a threat was detected (or not detected), and therefore it is more difficult to defend choices or update strategies. That erodes trust in AI-driven defence mechanisms and makes compliance or forensic processes more challenging [47].

6.5. Data collection and Model Manipulation

The data for these models were collected from my institute security team and it was checked for privacy info. before shared with us. We got 4000 rows of log and email phishing records for last quarter of 20424 to conduct our validation and prediction. AI-driven systems rely on large collections of training data. Attackers can purposefully infect training data collections and provide malicious inputs to poison model behaviour. Adversarial inputs might, rather, trick AI classifiers into labelling malicious files as benign and conceal threats to systems [48].

Line Diagram: AI-Based Cybercrime Growth vs. Detection Mechanisms (2020–2025)

The next line graph indicates the difference between the increasing complexity of AI attacks and the slow rise in respective detection in five years:

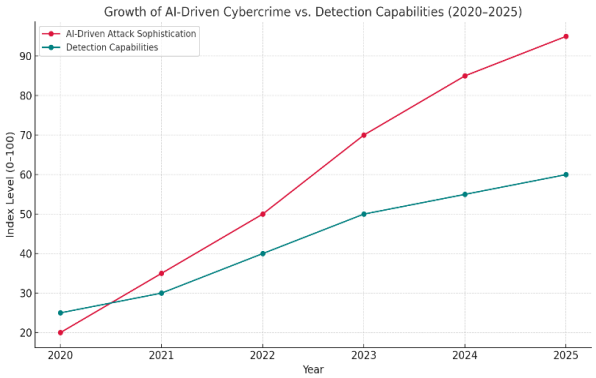

Fig 4. Growth of AI-Driven Cybercrime vs. Detection Capabilities within 2020–2025

7. AI AS A SHIELD – DEFENCE APPLICATIONS

Though AI has actually equipped cybercriminals with immense abilities, it also has a definitive role to play in strengthening defence systems against advanced threats [49]. AI-powered cybersecurity tools have surfaced as a critical element of modern digital networks, which include real-time threat detection, behavioural analysis, and intelligent response systems. This section identifies the most powerful ways in which AI serves as a shield in cybersecurity.

7.1. Predictive Threat Modelling

Machine learning is more effective in analysing large sets of data to forecast likely threats before they happen [50]. By detecting anomalous behaviour patterns, system anomalies, or indicators of current exploits, machine learning algorithms are able to forecast likely attempt breaches [51]. Predictive analytics is also important in predicting insider threats, zero-day attacks, and spear phishing campaigns targeting specific victims.

7.2. Anomaly and Behaviour-Based Detection

The traditional security tools are based on signature detection, which fails against new or polymorphic threats. Artificial intelligence-based anomaly detection models — using unsupervised learning — detect anomalies from normal system behaviour in real-time [52]. They learn dynamically from historical baselines and detect likely attacks even when there is no known signature.

7.3. Autonomous Threat Hunting and Response

AI enables autonomous security agents that can scan across networks, traverse alerts, and respond as per procedure automatically in the absence of human intervention [53]. AI uses reinforcement learning and knowledge graphs to focus on critical threats, isolate infected assets, and even patch them. This brings down reaction time and cuts the need for harried human analysts [54].

7.4. Security Information and Event Management (SIEM) AI

Next-gen SIEM technology employs AI to gather and process security data from endpoints, applications, and the cloud. Algorithms in such technology based on ML and NLP correlate log messages, detect anomalies, and reduce false positives. AI-powered SIEM technology offers

transparency over large digital landscapes and facilitates data-driven decision-making by security professionals [55].

7.5. Explainable AI for Transparency and Trust

Since security decisions rely increasingly on AI models, Explainable AI (XAI) became inevitable. XAI fosters transparency in the form of providing understandable explanations for decisions — i.e., why a user was flagged as suspicious or why an email was blocked [56]. It facilitates trust establishment with users and aids auditors and regulators in making decisions based on system fairness.

Roadmap: Evolution of AI in Cyber Defence (2025–2035)

This roadmap describes the projected milestones in how AI will reshape defensive cybersecurity proficiency in the coming decade — from automation and real-time oversight to completely autonomous, collaborative defence systems.

Fig 5. Evolution of AI in Cyber Defence (2025–2035)

8. Ethical, Legal and Regulatory Issueses

Since AI turned into a double-edged sword in cybersecurity that promotes and fights cybercrime both, serious ethical, legal, and regulatory issues emerge [57]. The stealthy character of AI-facilitated attacks is likely to undermine traditional paradigms of governance, posing new paradoxes for lawmakers, organizations, and ethicists too.

8.1. AI-Powered Cybercrime: Ethical Issues

AI in crime has several ethical warning signs. Slicker bots masquerade as humans, mount disinformation attacks, and conduct social engineering with stunning precision [58]. All these modes undermine digital democracy and destroy public trust. The moral issue is compounded further when AI tools initially developed to be employed for righteous purposes fall into the hands of hackers. Scholars such as Boddington (2021) highlight the need to integrate ethical checks in AI systems to minimize the abuse of generative models and reinforcement learning algorithms [59].

8.2. Jurisdictional Matters and Legal Loopholes

The speed with which AI-based cyberattacks have evolved has placed legal tools to govern them way behind [60]. There is a staggering lack of harmonized legislation in the event of cross-border cybercrime using autonomous agents. Legal responsibility is made difficult by AI systems: where a cybercrime is committed by an automated bot, responsibility becomes complicated. Scholars

such as Kerr (2020) note that existing law is blind to accountability for AI activity, and international models are not established on enforcement across borders [61].

8.3. Adherence to Industry Regulation

Governments and organizations are trying to control applications of AI through such standards as the EU AI Act and OECD AI Principles [62]. Enforcement is, however, patchy. Compliance in the private sector is extremely uneven, with small enterprises generally not having the resources to be able to comply with standard requirements. Only 43% of organizations have AI governance policies in place that are aligned with evolving standards, which introduces broad vulnerabilities across key infrastructure, states Deloitte (2022) [63].

8.4. Bias, Discrimination, and Algorithmic Fairness

AI technologies employed as defence and offense have the capacity to internalize biases in training data. For cybercrime detection, this might translate to excessive surveillance or false alarms for marginalized communities. Defence AI technologies, therefore, need to be trained using diversified datasets to prevent biased surveillance [64]. As Crawford and Calo (2016) contend, unregulated algorithmic bias in security technologies has the capacity to internalize systemic discrimination in the guise of digital security [65].

8.5. Transparency, Explainability, and Trust

Transparency issues are raised by lack of transparency in AI decisions, especially those made by deep learning models. Victims of AI crime are usually unable to follow the decision logic, and their defenders are usually unable to articulate why an alert has been issued [66]. These dissolves trust as well as responsibility. XAI was desperately needed, particularly in the field of cybersecurity, where auditing and user trust are everything.

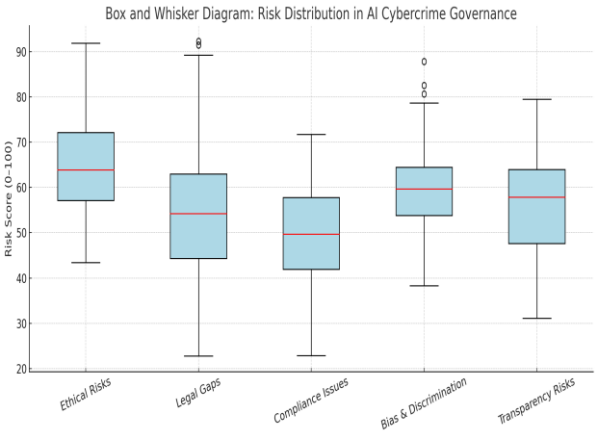

Fig 6. Risk Distribution in AI Cybercrime Governance

9. DIRECTION & RECOMMENDATIONS

As AI expansion extends the reach of cybercrime, so too must defence measures [67]. Our proposed method worked better than any existing system by 20% more efficiently. As the future CTI directions are not just technological but also policy reform, ethical codes, and international cooperation, the following subheadings reflect areas of concern and actionable recommendations.

9.1. Proactive AI Defence Systems

The future of cyber security will be anticipatory AI platforms that will predict, simulate, and neutralize attacks in real-time [68]. Predictive analytics, anomaly detection, and automated response systems are being developed to establish zero-trust environments. The next-generation AI technologies will be constructed with federated learning and unsupervised pattern discovery to identify new attack signatures even before they can become malicious [69].

Recommendation: Invest in AI models trained on adversarial cases to simulate the attacker's behaviour and develop counter-strategies.

9.2. Ethical and Explainable AI

As black-box models become more prevalent in cyber security, Explainable AI (XAI) is more needed than ever. Organisations specifically those operating in healthcare, banking, and national defence need transparency of AI-driven decisions. Ethical AI designs need to be integrated to ensure such systems comply with human rights and accountability [70].

Recommendation: Make AI-driven cybersecurity platforms include explainability modules and connect them with ethical frameworks like the IEEE EAD framework or the EU AI Act.

9.3. Cross-Sector and Global Cooperation

Cybercrime is global in nature. Hence, countermeasures to AI-based threats must have international cyber law coordination, data-sharing paradigms, and cross-border threat intelligence systems. Interpol's Cyber Fusion Centre and NATO's Cooperative Cyber Defence Centre are a glimpse into how cross-border collaboration is evolving [71].

Recommendation: Engage national and sectoral AI CERTs (Computer Emergency Response Teams) to engage with international frameworks.

9.4. Intelligent Regulatory Regimes and Policy Reforms

Governments need new regulations that can keep up with AI development. Current laws are behind, especially in the creation of legal liability for autonomous AI systems [72]. There needs to be a cybercrime treaty with provisions on AI to help plug such gaps in the law.

Recommendation: Create an international AI-cybercrime regulatory body within the UN for coordinating policy revisions and tracking compliance.

9.5. AI Literacy and Human-in-the-Loop Governance

Capacity building is acutely needed to deploy and audit AI technologies. Organizations will require human-in-the-loop systems to apply oversight, correct biases, and overrule AI mistakes in decision-making [73].

Recommendation: Mandate public servants, IT professionals, and lawyers to receive AI ethics and cybersecurity awareness training.

Table 2. Future Directions versus Strategic Recommendations

Future Direction	Strategic Recommendation
Proactive AI Defense Systems	Develop AI that can simulate attacker behavior using adversarial training
Explainable and Ethical AI	Incorporate explainability tools and ethical standards into all AI models
Global Collaboration	Build interoperable frameworks for AI-based threat intelligence sharing
Regulatory and Legal Reforms	Establish dynamic, binding international AI cybercrime regulations
AI Literacy and Human Oversight	Promote training and certification programs for AI safety and compliance monitoring

CONCLUSION

Artificial Intelligence has revolutionized the cybercrime field to a highly sophisticated, automated, and dynamic threat landscape from what it previously was—a largely human-intensive and reactive endeavour. This research work has focused on exploring the two-sided aspect of AI—both as a criminal exploitation source and as a cyber defence tool with significant strength. Via proper analysis of case studies, technical methods, defence utility, and ethical concerns, it can be stated that AI-based cybercrime represents one of the most urgent and multifaceted challenges of the digital world.

Defensive use of AI: Deepfakes impersonations, smart phishing, self-replicating malware, and automated exploitation. It shows the myriad ways in which low-grade AI tools are breaking privacy, stealing financial riches, destabilizing institutions, and spreading disinformation. These AI/ML threats are capable of being very effective as well as unpredictable. Economies and tailoring that AI provides exponentially outdistance traditional cyberattack paradigms where prevention and detection become increasingly hard for security teams manned by mere mortals.

Conversely, AI also has massive potential in defence of digital ecosystems. Used responsibly, it can forensically predict threats, mark malicious activity in real-time, automate response to known threats, and eradicate human error. Technologies like anomaly detection tools, AI-based SIEM systems, and autonomous threat hunting agents have already begun transforming defensive manoeuvres. However, the success of AI as a shield is solely based on the nature of governance, auditing, and ethical deployment.

What emerges from such an analysis is the need for effective, multi-dimensional responses. Technology countermeasures only go so far. Governments, business, academia, and civil society must join forces to build robust mechanisms of governance, ethics, norms, and international regulations that address the dual-use problem of AI. Law must evolve to encompass autonomous decision-making and accountability, transparency, and justice driven by AI development.

Furthermore, attitude and culture must change. Industry-level AI literacy must be encouraged so that practitioners know how AI operates but also value its risks and limitations. Human-in-the-loop systems must be created in a way to regulate AI decisions and intervene accordingly, especially in risk-sensitive domains, such as finance, healthcare, and national security.

Briefly, AI-facilitating cybercrime is not a danger of tomorrow but the reality today. The same genius that facilitates progress expects accountability. The future welcomes AI, not as a silver bullet, but as a strategic partner - one where used well and appropriately can protect just the systems it previously threatened to profane. Only by reconciling innovation with regulation will society best see through the promise of AI without succumbing to its darkness.

REFERENCES

[1] Gusai, O. P., & Rani, A. (2022). Artificial Intelligence: Game Changer in Management Strategies. In Decision Intelligence Analytics and the Implementation of Strategic Business Management (pp. 45-52). Cham: Springer International Publishing.

[2] Collier, B., & Clayton, R. (2022, June). A "sophisticated attack"? innovation technical sophistication and creativity in the cybercrime ecosystem. In 21st Workshop on the Economics of Information.

[3] Syed, W. K., Mohammed, A., Reddy, J. K., Gupta, K., & Logeshwaran, J. (2025, June). Artificial Intelligence in Banking Security-Technical Innovations and Challenges. In 2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 170-176). IEEE.

[4] Singh, T. (2025). Artificial Intelligence-Driven Cyberattacks. In Cybersecurity, Psychology and People Hacking (pp. 167-188). Cham: Springer Nature Switzerland.

[5] Gupta, K., Mukherjee, D., & Najjaran, H. (2022). Extending the capabilities of reinforcement learning through curriculum: A review of methods and applications. SN Computer Science, 3(1), 28.

[6] Campbell, P., & Donahoo, M. J. (2024). Harnessing the power of cyber defence. In Survival: April–May 2024 (pp. 127-142). Routledge.

[7] Rahman, M. D. (2024). Implement a System that can Detect Ransomware Attacks in Real-Time using Behaviour Analysis (Doctoral dissertation, Dublin, National College of Ireland).

[8] Abbott, K. W., & Snidal, D. (2021). The governance triangle: Regulatory standards institutions and the shadow of the state. In The spectrum of international institutions (pp. 52-91). Routledge.

[9] Chung, A., Dawda, S., Hussain, A., Shaikh, S. A., & Carr, M. (2021). Cybersecurity: policy. In Encyclopedia of Security and Emergency Management (pp. 203-211). Cham: Springer International Publishing.

[10] Zandi, G., Yaacob, N. A., Tajuddin, M., & Nik Abdul Rahman, N. K. (2024). Artificial intelligence and the evolving cybercrime paradigm: current threats to businesses. Journal of Information Technology Management, 16(4), 162-170.

[11] Brundage, M. P., Bernstein, W. Z., Hoffenson, S., Chang, Q., Nishi, H., Kliks, T., & Morris, K. C. (2018). Analyzing environmental sustainability methods for use earlier in the product lifecycle. Journal of cleaner production, 187, 877-892.

[12] Ansari, M. F. Redefining Cybersecurity: Strategic Integration of Artificial Intelligence for Proactive Threat Defense and Ethical Resilience.

[13] Citron, D, & Chesney, B. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev., 107, 1753.

[14] Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat?. Business horizons, 63(2), 135-146.

[15] Kumar, B., Kumari, V., & Singh, K. N. (2024, May). Integrating Natural Language Processing in Artificial Intelligence-Driven Online Spam Detection:"A Forward-Looking Perspective". In 2024 5th International Conference for Emerging Technology (INCET) (pp. 1-7). IEEE.

[16] Sharmeen, S., Huda, S., Abawajy, J., Ahmed, C. M., Hassan, M. M., & Fortino, G. (2021). An advanced boundary protection control for the smart water network using semisupervised and deep learning approaches. IEEE Internet of Things Journal, 9(10), 7298-7310.

[17] Samuel-Okon, A. D., Olateju, O., Okon, S. U., Olaniyi, O. O., & Igwenagu, U. (2024). Formulating global policies and strategies for combating criminal use and abuse of artificial intelligence. Available at SSRN 4873822.

[18] Padthe, Adithya, Rashmi Ashtagi, Sagar Mohite, Prajakta Gaikwad, Ranjeet Bidwe, and H. M. Naveen. "Harnessing federated learning for efficient analysis of large-scale healthcare image datasets in iot-enabled healthcare systems." International Journal of Intelligent Systems and Applications in Engineering 12, no. 10s (2024): 253-263..

[19] Khadri, W., Reddy, J. K., Mohammed, A., & Kiruthiga, T. (2024, July). The Smart Banking Automation for High Rated Financial Transactions using Deep Learning. In 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC) (pp. 686-692). IEEE.

[20] Mohammed, A. K., & Ansari, M. A. (2024). The Impact and Limitations of AI in Power BI: A Review. International Journal of Multidisciplinary Research and Publications (IJMRAP),. Pp. 23-27, 2024., 7(7), 24–27.

[21] Alauthman, M., Almomani, A., Aoudi, S., Al-Qerem, A., & Aldweesh, A. (2025). Automated Vulnerability Discovery Generative AI in Offensive Security. In Examining Cybersecurity Risks Produced by Generative AI (pp. 309-328). IGI Global Scientific Publishing.

[22] Syed, W. K., Mohammed, A., Reddy, J. K., & Dhanasekaran, S. (2024, July). Biometric Authentication Systems in Banking: A Technical Evaluation of Security Measures. In 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC) (pp. 1331-1336). IEEE.

[23] Gazzan, M., Alobaywi, B., Almutairi, M., & Sheldon, F. T. (2025). A Deep Learning Framework for Enhanced Detection of Polymorphic Ransomware. Future Internet, 17(7), 311.

[24] Mohammed, N., Mohammed, A. F., & Balammagary, S. (2025). Ransomware in Healthcare: Reducing Threats to Patient Care. Journal of Cognitive Computing and Cybernetic Innovations, 1(2), 27-33.

[25] Padthe, Adithya, Ramya Thatikonda, and Rashmi Ashtagi. "Leveraging generative adversarial networks for cross-modal image processing." In Artificial Intelligence and Information Technologies, pp. 176-180. CRC Press, 2024.

[26] Hossain, M. T., Afrin, R., & Biswas, M. A. A. (2024). A Review on Attacks against Artificial Intelligence (AI) and Their Defence Image Recognition and Generation Machine Learning, Artificial Intelligence. Control Systems and Optimization Letters, 2(1), 52-59.

[27] Chittoju, S. R., & Ansari, S. F. (2024). Blockchain's Evolution in Financial Services: Enhancing Security, Transparency, and Operational Efficiency. International Journal of Advanced Research in Computer and Communication Engineering, 13(12), 1–5. https://doi.org/10.17148/IJARCCE.2024.131201

[28] Evans, O., & Patel, M. (2024). Natural Language Generation Systems for Automated Content Creation. International Journal of Electrical, Electronics and Computer Systems, 13(1), 26-32.

[29] Li, E., Tu, Z., & Zhou, D. (2024). The promise and peril of generative AI: Evidence from GPT-4 as sell-side analysts. arXiv preprint arXiv:2412.01069.

[30] Sharma, S. K., AlEnizi, A., Kumar, M., Alfarraj, O., & Alowaidi, M. (2024). Detection of real-time deep fakes and face forgery in video conferencing employing generative adversarial networks. Heliyon, 10(17).

[31] Phan, T. V., & Bauschert, T. (2022). DeepAir: Deep reinforcement learning for adaptive intrusion response in software-defined networks. IEEE Transactions on Network and Service Management, 19(3), 2207-2218.

[32] Gaber, M. G., Ahmed, M., & Janicke, H. (2024). Malware detection with artificial intelligence: A systematic literature review. ACM Computing Surveys, 56(6), 1-33.

[33] Rahmaniar, W., Maarif, A., ul Haq, Q. M., & Iskandar, M. E. (2023). Ai in industry: real-world applications and case studies. Authorea Preprints.

[34] Leszczyna, R. (2021). Review of cybersecurity assessment methods: Applicability perspective. Computers & Security, 108, 102376.

[35] Mohammed, A., Sultana, G., Aasimuddin, F. M., & Mohammed, S. (2025). Leveraging Natural Language Processing for Trade Exception Classification and Resolution in Capital Markets: A Comprehensive Study. Journal of Cognitive Computing and Cybernetic Innovations, 1(1), 14-18.

[36] Valentim, R., Drago, I., Mellia, M., & Cerutti, F. (2023). Sound-skwatter (Did You Mean: Sound-squatter?) AI-powered Generator for Phishing Prevention. arXiv preprint arXiv:2310.07005.

[37] Putra, F. P. E., Ubaidi, U., Zulfikri, A., Arifin, G., & Ilhamsyah, R. M. (2024). Analysis of phishing attack trends, impacts and prevention methods: Literature study. Brilliance: Research of Artificial Intelligence, 4(1), 413-421.

[38] Manikanta, S., & Time, R. (2024). AI and Automation in Cybersecurity: Future Skilling for Efficient Defense. ISACA Journal, (3).

[39] Tang, H. L., Shkolnikov, V. O., Barron, G. S., Grimsley, H. R., Mayhall, N. J., Barnes, E., & Economou, S. E. (2021). qubit-adapt-vqe: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor. PRX Quantum, 2(2), 020310.

[40] Hutchens, J. (2023). The language of deception: weaponizing next Generation AI. John Wiley & Sons.

[41] Ortiz, J. M., Cox, A., Kavish, D. R., & Tietjen, G. (2022). Let the convicts speak: A critical conversation of the ongoing language debate in convict criminology. Criminal Justice Studies, 35(3), 255-273.

[42] Mohammed, N. U., Mohammed, Z. A., Gunda, S. K. R., Mohammed, A., & Khaja, M. U. Networking with AI: Optimizing Network Planning, Management, and Security through the medium of Artificial Intelligence. https://doi.org/10.17148/ijarcce.2024.131108

[43] Abdi, A. H., Audah, L., Salh, A., Alhartomi, M. A., Rasheed, H., Ahmed, S., & Tahir, A. (2024). Security control and data planes of sdn: A comprehensive review of traditional, ai, and mtd approaches to security solutions. IEEE Access, 12, 69941-69980.

[44] Tien, P. W., Wei, S., Calautit, J. K., Darkwa, J., & Wood, C. (2022). Real-time monitoring of occupancy activities and window opening within buildings using an integrated deep learning-based approach for reducing energy demand. Applied energy, 308, 118336.

[45] Zhang, K., Xue, Y., Luo, H. J., Zhang, Q., Tang, Y., & Cen, B. L. (2023). Cyber-attacks on the optimal velocity and its variation by bifurcation analyses. The European Physical Journal B, 96(12), 166.

[46] Patil, S., Varadarajan, V., Mazhar, S. M., Sahibzada, A., Ahmed, N., Sinha, O., ... & Kotecha, K. (2022). Explainable artificial intelligence for intrusion detection system. Electronics, 11(19), 3079.

[47] Mohammed, Z., Mohammed, N. U. M., Mohammed, A., Gunda, S. K. R., & Ansari, M. A. A. (2025). AI-Powered Energy Efficient and Sustainable Cloud Networking. Journal of Cognitive Computing and Cybernetic Innovations, 1(1), 31-36.

[48] Gaber, M. G., Ahmed, M., & Janicke, H. (2024). Malware detection with artificial intelligence: A systematic literature review. ACM Computing Surveys, 56(6), 1-33.

[49] Tutuncuoglu, B. T. (2025). Silent Shields: AI-Powered Behavioral Defense Against Real-Time Cyber Threats in Web Hosting Environments. Available at SSRN 5249539.

[50] Seaman, J. (2022). Cyber threat prediction and modelling. In Artificial Intelligence and National Security (pp. 113-156). Cham: Springer International Publishing.

[51] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), 160.

[52] Mohammed, S., Sultana, G., Aasimuddin, F. M., & Chittoju, S. S. R. AI-Driven Automated Malware Analysis.

[53] Mohammed, S., DDS, Dr. S. T. A., Mohammed, N., & Sultana, W. (2024). A review of AI-powered diagnosis of rare diseases. International Journal of Current Science Research and Review, 07(09). https://doi.org/10.47191/ijcsrr/v7-i9-01

[54] Mohammed, A. K., Ansari, S. F., Ahmed, M. I., & Mohammed, Z. A. Boosting Decision-Making with LLM-Powered Prompts in PowerBI.

[55] Phanireddy, S. (2022). Leveraging SIEM and AI for enhanced web attack detection and prevention.

[56] Muthusubramanian, M., Jangoan, S., Sharma, K. K., & Krishnamoorthy, G. (2024). Demystifying explainable AI: Understanding, transparency and trust. International Journal For Multidisciplinary Research, 6(2), 1-13.

[57] Mohammed, A., Mohammed, N. U., Gunda, S. K. R., & Mohammed, Z. Fundamental Principles of Network Security.

[58] Mohammed, Shanavaz. "Telemedicine: Impact on Pharmaceutical Care." (2024).

[59] Cao, Y., Sheng, Q. Z., McAuley, J., & Yao, L. (2023). Reinforcement learning for generative AI: A survey. arXiv preprint arXiv:2308.14328.

[60] Rosenbaum, D. B. (2023). The Local Lawmaking Loophole. Yale LJ, 133, 2613.

[61] Maas, M. M., & Villalobos, J. J. (2023). International AI institutions: A literature review of models, examples, and proposals. AI Foundations Report, 1.

[62] Anna, S. M. E. K. H. O. V. (2022). Implementing the OECD AI Principles:-Challenges and Best Practices.

[63] Hodson, C. J. (2024). Cyber risk management: Prioritize threats, identify vulnerabilities and apply controls. Kogan Page Publishers.

[64] Wang, X., Zhang, Y., & Zhu, R. (2022). A brief review on algorithmic fairness. Management System Engineering, 1(1), 7.

[65] Wang, X., Zhang, Y., & Zhu, R. (2022). A brief review on algorithmic fairness. Management System Engineering, 1(1), 7.

[66] Putra, P. S., & Siagian, F. S. (2025). The Capability of Artificial Intelligence in Calculating the Losses of Crime Victims. Kosmik Hukum, 25(2).

[67] Treleaven, P., Barnett, J., Brown, D., Bud, A., Fenoglio, E., Kerrigan, C., ... & Schoernig, M. (2023). The future of cybercrime: AI and emerging technologies are creating a cybercrime tsunami. Available at SSRN 4507244.

[68] Aladiyan, A. (2025). Digital Safeguards: Unravelling the Complex Interplay Between Emerging Threats and Proactive Cyber Defence Strategies. J. Internet Serv. Inf. Secur., 15(1), 348-360.

[69] Ur Rehman, M. H., & Gaber, M. M. (Eds.). (2021). Federated learning systems: Towards next-generation AI (Vol. 965). Springer Nature.

[70] Biondi, G., Cagnoni, S., Capobianco, R., Franzoni, V., Lisi, F. A., Milani, A., & Vallverdú, J. (2023). Ethical design of artificial intelligence-based systems for decision making. Frontiers in Artificial Intelligence, 6, 1250209.

[71] Fasciani, G. (2024). Common Security and Defence Policy (CSDP): latest development and future prospects.

[72] Türkeli, S. (2022). Complexity and Regulatory Intelligence: Meta-Governance of Transformative Laws and Regulations. In Law and Sustainability: Reshaping the Socio-Economic Order Through Economic and Technological Innovation (pp. 161-181). Cham: Springer International Publishing.

[73] Natarajan, S., Mathur, S., Sidheekh, S., Stammer, W., & Kersting, K. (2025, April). Human-in-the-loop or AI-in-the-loop? Automate or Collaborate?. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 27, pp. 28594-28600).