

Policy vs. Practice: A Comparative Analysis of EU and US Responses to AI-Driven Cyber Incidents

Laura Rueda García

Department of Information Technology
Illinois Institute of Technology
Chicago, IL
lruedagarcaa@hawk.illinoistech.edu

Felipe Susaeta Miguel

Department of Information Technology
Illinois Institute of Technology
Chicago, IL
fsusaetamiguel@hawk.illinoistech.edu

Abstract—Previous research, like the 2024 study by Ijaiya and Odumuwaun, has compared the different rules for AI in the European Union (EU) and the United States (US). We know the EU has one main set of strict rules, while the US has many separate ones. However, we don't know how well these rules actually work against new cyberattacks that use AI, like deepfakes or advanced phishing emails. This paper investigates that question by looking at real-world incidents. We manually searched for and analyzed a handful of publicly known AI-driven cyberattacks in both the US and the EU. For each case, we examined how the attack happened, how the company reacted, and what government agencies did in response.

Index Terms—Cybersecurity, Artificial Intelligence, AI Governance, Risk Management, GRC.

I. INTRODUCTION

A. Context

Artificial intelligence is reshaping the cybersecurity world by enabling more effective defenses and more difficult attacks. AI enables a new generation of social engineering where adversaries use large language models and synthetic media to create highly convincing, context-aware lures that bypass traditional filters and human skepticism. In recent years, AI-powered phishing attacks have increased by 125% [1] and enterprise risk executives now list AI-enhanced attacks among their top emerging risks, reflecting that the democratization of generative tools increases the scale, speed, and sophistication of phishing, impersonation, and automated reconnaissance campaigns. [2] Because of threats like this, governments and organizations are developing new frameworks addressing data protection, transparency, and accountability of the new AI driven world, such as the European General Data Protection Regulation and the EU Artificial Intelligence Act or the NIST Framework in the US.

B. Problem Statement

Existing comparative work describes a clear regulatory divergence: the European Union relies on a unified, rights-based model with the GDPR and the EU AI Act while the United States depends on a fragmented patchwork of sectoral laws, state privacy statutes and voluntary federal guidance. But, What remains unanswered, is how these differing regulatory regimes perform when confronted with AI-driven cyber-threats. So, in this study we want to address this question.

C. Paper Structure

The paper has 4 main sections:

- **Introduction:** This is where it is explained what is inside the paper, our motivations and the main problem we want to address.
- **Literature Review:** This is where the analysis of the previous work and some important information about the cyber-security world are.
- **Proposed Methodology:** This section explains how the investigation was conducted, explaining the methods used, how we gather the information and difficulties we found regarding this study.
- **Analysis and discussion:** This is where the main problem of the paper is addressed and where the results of our investigation are. This is the biggest section of the paper and we provide numerous subsections to guide the reader through it.
- **Conclusion and future work:** Here, we give our insights and explore solutions to the problems discovered in the sections before. Also, we will look into the future to predict how the AI will shape these years to come.
- **References:** The sources of the information used in the study.

II. LITERATURE REVIEW

A. Comparison of frameworks

In this paper we will look into the frameworks used in the United States and the European Union. These frameworks give us directions, good practices and helps the organizations work safe with AI tools.

1. European Union

The European Union has two main frameworks that focus in cybersecurity and AI. The first one is called **General Data Protection Regulation (GDPR)** is a law introduced in 2018 and its main focus is in protecting data from users and limiting the data that organizations can use and collect from their users. The main principles are lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality. People are granted rights to access their data, change it, erase it, limiting their processing, etc. This framework is extremely oriented

in user data protection and establish heavy fines in case an organization infringes this law.

But, after the rise in popularity in AI, in 2021 the EU introduced the first framework regarding the use of AI. The **EU Artificial Intelligence Act (AI Act)** aims to promote the safe and trustworthy development of AI systems while keeping the usual security measures the EU has for the data from their citizens. In this law, the EU ranks each AI system on a danger level regarding the risk the system has on the population. It goes from minimal risk (video games or filters) to unacceptable risk (social rankings) and for each level it has different security requirement the system must have. Also, systems has to have transparency facilitating human oversight. These 2 laws are supported by other frameworks like ePrivacy Directive or the EU Cybersecurity Act.

The EU Cybersecurity Act(2019) is a permanent mandate for ENISA(European Union Agency for Cybersecurity) and establishes a European cybersecurity certification framework that enables the development of common, risk-based certification schemes for ICT products, services, and processes, including components and lifecycles of AI systems.

2. US

The US is a little more lacking regarding in regulation compared to the EU. It has no unified framework, it relies in a mix of sector directives, state laws and federal guides and present challenges for AI systems due to inconsistent standards, overlapping requirements, and gaps in coverage. It has some laws like the Gramm-LeachBliley Act (GLBA) and the California Consumer Privacy Act (CCPA) in which data privacy and protection is discussed. But recently in 2023, the federal government released the NIST AI Risk Management Framework that is a optional framework to promote trustworthy AI systems declaring their interest in regulating AI but as indicated before, this is an optional framework, leaving the final decision to implement this to the organization so it does not reach to every organization. In a state perspective, California with its CCPA leads the way in regulation in the US, giving rights to their citizens like the ones we saw in the EU. People can ask for their data, ask to erase it, change it... It affect AI systems since it also establish rules on how these systems can collect information and use it. It also establish a mandatory algorithmic impact assessments for high-risk AI systems like credit scoring systems.

Regarding US Cybersecurity policies, the NIST Framework provides guidelines for securing AI systems against cyber threats and its organized around five core functions: Identify, Protect, Detect, Respond, and Recover. This framework is also an optional guideline that has no legal consequences

3. Comparison

The EU regulation is extremely harsh on organization regarding on the privacy and security of the users information with unified and strong frameworks. On the contrary, the US relies on optional guidelines and sector specific directives that incentivizes innovation but this lack of strict guidelines presents a security risk if the enterprise has not made an investment. Practically, the EU approach is more likely to

drive standardized certification and uniform evidentiary requirements that aids incident response and forensic readiness, and the NIST approach is more likely to accelerate flexible, sector-driven operational improvements and consensus on metrics.

B. A classification of AI-driven CyberAttacks

AI is a double-edged sword for cybersecurity. It can be used to defend against attacks, but it can also be used to create novel, more powerful ones. Researchers and government agencies have identified two primary categories of these new threats: attacks that target AI systems directly, and attacks that leverage AI as a weapon.

1. Attacks directed towards AI systems

These attacks are designed to deceive, corrupt, or compromise the integrity of AI systems that companies use for their security and daily operations. Given the ubiquity of AI in contemporary society, in everything from medical diagnostics to financial decision-making, these attacks are particularly dangerous.

- Adversarial Examples:

This is a sophisticated method of deceiving an AI system by making tiny, frequently imperceptible alterations to data. According to a report by the Swedish Defence Research Agency (FOI) [3] the modification of a few number of pixels in a photograph has the potential to deceive facial recognition systems, leading to the misidentification of a person, or to cause a self-driving car to misinterpret a stop sign. According to Javaid et al. (2021) [4], this phenomenon constitutes a big problem, as it has the potential to compromise the reliability of security systems that use AI. An attacker could leverage this technique to disguise a piece of malware, making an AI-powered antivirus misidentify a hazardous virus as a harmless program.

- Data and Model Poisoning:

This category of cyberattack is initiated at the system's initial training phase, when the AI system is acquiring knowledge from data. This phenomenon can be conceptualized as a form of contamination of the "textbooks" the AI studies. The FOI report describes how attackers can furtively insert malicious data during the training of the AI. This can result in a complete breakdown of the AI's capacity to make optimal decisions. Javaid et al. (2021) point out that a "poisoned" AI might be trained to always permits the entry of a specific hacker into the system, creating a cover backdoor that is subsequently difficult to detect.

- Model Inversion and Membership Inference:

The objective of these attacks is to extract private information from the AI model itself. As Pandey and Pandey (2024) [5] illustrate, hackers can pose questions to the AI and meticulously analyze the responses to deduce the confidential information that the AI was trained on. This phenomenon is referred to as "model inversion". For instance, they could potentially reconstruct faces from

a facial recognition model. A similar attack, known as membership inference, is employed to determine if a particular individual's information was included in the AI's training dataset. This is a major privacy violation because it could reveal if a person's private medical records were used to train a health-related AI.

2. Attacks powered by AI Systems

This category includes attacks where criminals leverage AI to amplify the scale, sophistication and efficacy of their attacks. This allows them to launch more sophisticated attacks with reduced effort.

- **AI-Enhanced Social Engineering:**

According to researchers in the field, this constitutes one of the most significant contemporary threats. The FOI report highlights the potential of deepfakes (highly realistic synthetic videos or audio) to impersonate prominent figures, such as a CEO, with the intention of deceiving employees into transferring funds. Unlike old phishing emails that were covered with spelling errors, a review by Thuraisingham et al. (2022) [6] has noted the emergence of generative AI as a tool to generate thousands of unique, convincing, and personalized phishing emails on a massive scale. This results in a significant increase in the difficulty for both individuals and security software to detect them.

- **AI-Powered Malware and Automation:**

Criminals are leveraging AI to develop more sophisticated malware. Pandey and Pandey (2024) have noted that this "intelligent" malware has the capacity to modify its own code, in order to evade antivirus software, thereby becoming a "moving target" that is extremely difficult to capture. This represents a substantial advancement from earlier viruses, which possessed a fixed signature. Additionally, AI is employed to automate preliminary phases of an attack. Thuraisingham et al. (2022) describe a method for rapidly scanning thousands of systems across the internet using AI to identify vulnerabilities. In addition, Javaid et al. (2021) state that AI assists attackers in automatically identifying and acquiring knowledge about valuable targets, enabling them to plan their attacks more efficiently and quickly than a human could.

III. PROPOSED METHODOLOGY

The central hypothesis of this research is that the European Union's (EU) stricter, unified regulations lead to a more effective response to AI-driven cyber incidents compared to the fragmented framework in the United States (US). To test this hypothesis, the present study compares the cybersecurity and AI regulations of both regions by examining how each one handles real AI-driven cyberattacks. This study functions as a practical sequel to the baseline paper, which exclusively looked at the different regulations. The objective of this study is to assess the efficacy of these rules in real world settings.

Initially, the research attempted to find reports of these attacks automatically using a Python script. The objective of this methodology was to meticulously scan government

websites and technical security blogs for any mention of AI-driven incidents. This automated approach, however, was ultimately found to be ineffective. The script returned a high volume of irrelevant information, such as marketing pages and general policy documents, instead of specific reports on cyber incidents. This initial failure became an important discovery in itself. It demonstrates that even though AI threats are widely discussed, as noted by Cyber Defense Magazine [7], which notes that while AI threats are a subject of extensive discussion, very few specific, real-world incidents are publicly documented in a way that is easy to find. This phenomenon is particularly strange when other research from institutions such as MIT suggests that AI is already a component in numerous cyberattacks, such as ransomware [8]. This suggests a significant discrepancy between the perceived threat and public reporting, suggesting that AI-driven attacks are likely underreported or their technical details are intentionally not shared. This situation makes a manual, detailed search for case studies essential.

As a result, the research strategy underwent a shift towards a more manual approach, entailing manually finding and analyzing a smaller number of high-quality, verifiable case studies. The new process is composed of three main steps:

- **Finding Cases:** The initial step entails conducting a manual search for recent, specific incidents in both the US and EU where AI served as a fundamental component of the attack. This approach prioritizes the quality and relevance of each case study over the quantity of data.
- **An investigation into the credibility of the sources is imperative.** The next step, entails a meticulous examination of the sources. Information is gathered from reputable technical news and official reports to confirm the details of each case. This methodological approach ensures that the analysis is grounded in solid and verifiable evidence, enhancing the reliability of the research findings.
- **Analyzing the Cases:** The third step is the analysis of the cases. A consistent checklist is applied to each case study to examine the incident, the response from the company, and, most importantly, the response from regulatory and governmental bodies. The analysis looks into whether law enforcement was involved, if specific laws (like GDPR) were applied, and if any fines or official recommendations were made.

By examining these real-world cases, this method facilitates a direct comparison of how the EU and US apply their AI regulations in practice. This methodological approach enables the empirical evaluation of the theoretical framework outlined in the baseline paper.

IV. ANALYSIS AND DISCUSSION

V. CONCLUSION AND FUTURE WORK

REFERENCES

- [1] W. P., "Cross-border ai regulatory strategies," AI Regulation Studies, May 2023.

- [2] G. P. Release, "Gartner survey shows ai-enhanced malicious attacks are a new top emerging risk for enterprises," Gartner, May 2024. [Online]. Available: <https://www.gartner.com/en/newsroom/press-releases/2024-05-22-gartner-survey-shows-ai-enhanced-malicious-attacks-are-new0>
- [3] P. Johnson, S. Ek, and M. Arvidsson, "Adversarial artificial intelligence: A study of the threat of adversarial machine learning and the impact on the swedish armed forces," Swedish Defence Research Agency (FOI), Tech. Rep. FOI-R-4971-SE, May 2020.
- [4] M. U. Javaid, S. Ali, A. Khalid, M. Asim, U. Ullah, and J. B. Sarfraz, "Cyber threats and security challenges in the era of artificial intelligence and machine learning," *Procedia Computer Science*, vol. 184, pp. 822–828, 2021.
- [5] O. Pandey and J. Pandey, "Cybersecurity in the age of emerging technologies: Challenges and solutions," in *Intelligent Systems and Applications*. Springer Nature Switzerland, 2024, pp. 179–185.
- [6] B. Thuraisingham, L. Khan, M. A. Al-Faruq, and M. Kantarcioglu, "Artificial intelligence and cybersecurity: a systematic literature review," *Applied Artificial Intelligence*, vol. 36, no. 1, 2022.
- [7] P. Paganini, "The growing threat of ai-powered cyberattacks in 2025," *Cyber Defense Magazine*, January 2025. [Online]. Available: <https://www.cyberdefensemagazine.com/the-growing-threat-of-ai-powered-cyberattacks-in-2025/>
- [8] C. at MIT Sloan, "80% of ransomware attacks now use artificial intelligence," MIT Sloan Ideas Made to Matter, May 2024. [Online]. Available: <https://mitsloan.mit.edu/ideas-made-to-matter/80-ransomware-attacks-now-use-artificial-intelligence>