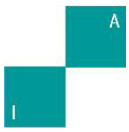
Inteligencia Artificial, 27(73) (2024), 38-54
doi: 10.4114/intartif.vol27iss73pp38-54



INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Ethics in Artificial Intelligence: an Approach to Cybersecurity

Ariel López González[1], Mailyn Moreno-Espino[2], Ariadna Claudia Moreno Román[1], Yahima Hadfeg Fernández[3], Nayma Cepero Pérez[4]

1 Escuela Superior de Ingeniería Mecánica y Eléctrica, Unidad Culhuacán, Instituto Politécnico Nacional, Ciudad de México 04440, México; lopez.gonzalez.ariel1@gmail.com, moreno.roman.ariadna.claudia@gmail.com
2 Centro de Investigación en Computación, Instituto Politécnico Nacional, Ciudad de México 07700, México; mmorenoe2022@cic.ipn.mx
3 Facultad de Economía y Administración, Universidad Católica del Norte, Antofagasta 0610, Chile; yahima.hadfeg01@ucn.cl
4 Facultad de Ingeniería Informática, Universidad Tecnológica de La Habana "José Antonio Echeverría", cujae, La Habana 19390, Cuba; ncepero@ceis.cujae.edu.cu

Abstract In the paper, an analysis is conducted on the intricate relationship between ethics, artificial intelligence, and cybersecurity. The ethical principles that govern the advancement of AI are examined, alongside the security issues that arise from its implementation. The ethical utilization of artificial intelligence in the realms of cybersecurity and hacking is explored. Emphasis is placed on the significance of AI ethics, particularly in terms of transparency, accountability, and fairness. Additionally, the paper delves into the security challenges that emerge as AI is adopted, such as safeguarding user privacy and ensuring equitable access to the technology.

Keywords: Ethics, Artificial Intelligence, Cybersecurity, Ethical hacking.

1 Introduction

Ethics is a word with many meanings and facets. Ethics is a field of study that has evolved throughout history and has been addressed by numerous philosophers, thinkers, and cultures at different times. Ethics has no single origin or definition, as it has been the subject of reflection and debate for millennia. Ethics has its roots in ancient Greek philosophy. Greek philosophers such as Socrates, Plato, and Aristotle made significant contributions to the study of ethics [39, 59]. In the modern era, during the Enlightenment of the 18th century, philosophers like Immanuel Kant developed ethical theories based on reason and moral duty. Kant proposed deontological ethics, which focuses on duty and morality based on universal principles [45]. In the contemporary era, namely the 20th century, philosophers like Jean-Paul Sartre and Albert Camus explored existentialist ethics, which centers on individual responsibility and freedom [65]. New branches of ethics also emerged, such as professional ethics and bioethics, which address specific ethical issues in fields like medicine, robotics, or Artificial Intelligence.

Artificial Intelligence is a field that focuses on creating systems and machines capable of performing tasks that typically require human intelligence. The challenges of AI include developing ethical systems, transparency in machine decisions, and understanding how AI can impact society [10, 16]. Since the emergence of Artificial Intelligence and the revolutionary advancements it has brought, there has been a stir ranging from people who fear artificial intelligence to those who incorporate it imperceptibly into

ISSN: 1137-3601 (print), 1988-3064 (on-line)

their daily lives. With its ability to learn from data and make decisions, AI poses a series of ethical and practical challenges [27, 55]. On one hand, Artificial Intelligence promises innovative solutions in areas such as medicine, industry, education, and more. It can improve efficiency, automate tasks, and provide new opportunities. However, its adoption raises fundamental ethical questions, such as data privacy, algorithmic discrimination, and biased decision-making [50, 25, 19].

Among the fields of computer science, Artificial Intelligence is a fundamental area. Nevertheless, there are other highly relevant and current areas, such as Cybersecurity. Cybersecurity refers to the practice of protecting computer systems, networks, and data against cyber threats such as hacker attacks, malware, and information theft [21]. Challenges in cybersecurity include maintaining the confidentiality, integrity, and availability of information and protecting user privacy. With the growing dependence on technology, cybersecurity has become a critical challenge in the digital age [46]. In the context of cybersecurity, protection against cyber threats becomes essential. The interconnection of devices and online systems creates vulnerabilities that cybercriminals can exploit [33, 14]. Cybersecurity seeks to ensure that our systems and data are safe from attacks but also presents challenges in terms of protecting privacy and ensuring equity in access to technology [52, 44].

The development of Artificial Intelligence has enabled the creation of techniques to mitigate security risks in computer systems [44]. However, just as Artificial Intelligence has allowed the development of new methods to enhance cybersecurity, its unethical use also poses a potential tool for cybercriminals and cyberattacks [43, 83]. This article explores the intersection of ethics, artificial intelligence, and cybersecurity, analyzing the ethical principles relevant to Artificial Intelligence and the security challenges associated with its use. Furthermore, it considers the impact of these issues on society and how ethics and ethical decision-making play a fundamental role in this context.

2 Artificial Intelligence, Ethics and Cybersecurity

2.1 Ethical Principles of Artificial Intelligence

Whenever a branch of science emerges that has such a significant impact on people's lives, ethical principles arise to be respected in the work of that field. Artificial Intelligence is no exception to this; in Artificial Intelligence, ethics plays a fundamental role in formulating guidelines and principles that guide the development and implementation of systems using AI techniques.

Ethical principles in AI have been proposed and their formulations continue to improve over the years, and these principles are essential to ensure that AI is used responsibly. There are many ethical principles that have been formulated in the context of AI; we present some that are highly relevant [20, 25, 19, 53, 12, 32, 37, 82, 78, 6, 1]:

- Transparency and Explainability: One of the fundamental principles is transparency in AI decision-making processes. AI systems must be able to explain how they arrive at their conclusions, allowing users to understand and trust their decisions. This is especially important in critical applications, such as healthcare and legal decision-making.
- Fairness and Non-Discrimination: AI should not perpetuate or amplify biases or discrimination existing in society. AI systems should be trained with diverse and representative data to ensure that there are no gender, race, or other characteristic biases in the outcomes. Measures should be implemented to correct and prevent discrimination.
- Data Privacy: Data privacy protection is essential. AI systems must handle data securely and respect individuals' privacy. This involves compliance with regulations such as the General Data Protection Regulation (GDPR) in the European Union.
- Accountability and Responsibility: Responsibility for decisions made by AI systems must be established. This includes identifying who is accountable in case of negative outcomes or harm caused by AI. Clarity in responsibility is crucial for addressing legal and ethical issues.
- Beneficence and Non-Maleficence: AI systems should seek the benefit of humanity and avoid causing harm. This implies that AI developers and users must carefully consider the ethical implications of its use and take measures to minimize risks.

- Collaboration and Human Oversight: AI should not completely replace human supervision and decision-making. Instead, it should be used as a collaborative tool that enhances human decision-making, especially in critical areas such as medicine and security.
- Secure Development: Cybersecurity is essential in AI system development. Systems must be protected against attacks and vulnerabilities to prevent potential negative consequences.
- Just Distribution of Benefits: Ensuring that the benefits of AI are fairly distributed in society and not concentrated in a few is necessary. This involves considering aspects such as equitable access to technology and training.

2.2 Cybersecurity in the context of Artificial Intelligence

Cybersecurity is one of the branches in which AI is currently being applied, and one which includes many ethical definitions [76]. In an increasingly digitized world, information protection and online security have become key priorities. Cybersecurity refers to the measures and practices adopted to protect computer systems and networks against cyberthreats, such as data theft, malware and hacker attacks [34].

The main mechanisms used in cybersecurity are intrusion detection systems, firewalls, data loss prevention systems and antivirus software. These traditional tools are essential for detecting and mitigating known risks, but their ability to cope with constantly evolving threats is limited [34].

Nowadays, cybercriminals are using more sophisticated tactics and increasingly targeted attacks [80]. Traditional cybersecurity mechanisms rely on predefined rules and signatures to detect and block known threats. However, these approaches are static and cannot rapidly adapt to new forms of cyber-attack. What's more, cybercriminals are also using sophisticated techniques to bypass the rules and avoid detection by traditional systems [73].

This is where AI can make the difference [11]. In cybersecurity, AI offers more advanced detection and response capabilities through the use of machine learning algorithms and data analysis techniques. AI can analyze large volumes of data in real time and identify patterns, anomalies and malicious behavior that might not be detected by traditional systems. This enables earlier detection and faster response to cyberthreats, reducing detection time and potential damage [76].

In addition, AI can adapt and learn continuously, improving its ability to identify and predict attacks as they evolve. It can thus keep abreast of the latest tactics and techniques used by cybercriminals, and provide more effective defenses against emerging threats [76].

Among the AI algorithms most commonly used in cybersecurity are neural networks, anomaly detection algorithms, Bayes classifiers and genetic algorithms. These algorithms can analyze large quantities of security data, identify suspicious patterns and make automated decisions based on the analysis of this data [80]. In addition, intelligent agents - software entities capable of sensing the environment, making decisions and acting autonomously - play a crucial role in cybersecurity. These intelligent agents interact with other security systems and coordinate responses to potential threats, using artificial intelligence algorithms to analyze data, identify malicious behavior and take action to protect systems against cyber-attacks. Their ability to adapt and learn continuously makes them a valuable tool for mitigating risks and protecting IT systems and networks [54].

However, the use of AI in cybersecurity also raises important ethical issues [85, 75]. One of these challenges concerns the interpretability and transparency of the results of AI algorithms. In many cases, AI systems make decisions based on complex models that are difficult to understand, which can lead to a lack of trust and difficulty in explaining and justifying the decisions made [34, 54].

In addition, there are ethical challenges linked to privacy and data security. The implementation of artificial intelligence systems in the field of cybersecurity involves the collection and processing of large quantities of sensitive data. This raises concerns about user privacy and the appropriate use of such data. It is essential to put in place robust data protection mechanisms and ensure that the ethical principles of privacy and confidentiality are respected [48].

Another ethical issue linked to the use of artificial intelligence in cybersecurity is the risk of discrimination and bias in algorithms. If AI models are trained on biased or discriminatory datasets, they risk reproducing and reinforcing these biases in their decisions. This can lead to unfair discrimination against certain groups, or to unintended exclusions [73, 54]. There are also issues around balancing job

opportunities and automation, and the accountability and ownership of decisions made by AI models [34, 51].

In summary, AI has revolutionized cybersecurity by enabling more effective detection and faster response to cyberthreats. However, it is essential to address the ethical challenges associated with the use of AI in this field. It is necessary to guarantee the transparency and interpretability of algorithms, protect privacy and data security, and avoid discrimination and bias in AI systems used in cybersecurity. This will enable us to fully exploit the benefits of AI to protect our systems and networks without compromising fundamental ethical values.

3 Security challenges associated with the use of artificial intelligence

Security challenges in AI are a fundamental concern in the development of this advanced technology. Among the most concerning is the interpretability and transparency of the results obtained from the models. As AI algorithms become more complex and deeper, their operation becomes less clear to humans, raising significant safety and ethical issues [70]. The lack of interpretability and transparency in AI models presents several problems [34]:

- Low trust: When users or experts cannot understand how an AI model arrives at its conclusions, it is difficult to fully trust its results. This is especially critical in applications where important decisions are made, such as in healthcare or legal decision-making.
- Difficulties in bias detection: If it is not possible to understand how an AI model makes decisions, it is also difficult to identify and address possible biases and discrimination in its results. AI models can learn biases from the data they are trained on and perpetuate them without our awareness.
- Security and adversarial attacks: The lack of interpretability can be exploited by attackers to exploit vulnerabilities in the AI model. Attackers can find ways to manipulate the algorithm's operation without being detected, which can have serious consequences in applications such as cybersecurity or fraud detection.

To address this challenge, AI researchers and developers are working on interpretability and explainability techniques [63]. These techniques seek to make AI models more transparent and understandable to humans by providing explanations for how certain decisions or predictions were made. Some of these techniques include using simpler and more understandable models, visualizing the internal activity of the model, and generating explanations based on rules or examples [17].

Another significant challenge is data privacy and security. AI requires large amounts of data for training and operation, which can lead to various vulnerabilities and risks to user privacy and information security. Some key aspects of this challenge are explained below [2, 81]:

- Data leaks: Collecting, storing and processing data to train AI models can lead to potential data leaks. If data is not handled properly or is not protected with robust security measures, it could become accessible to unauthorized individuals or fall victim to cyber attacks, compromising the privacy of individuals and organizations.
- Sensitive and personal data: Data used in training AI models often contains personal and sensitive information, such as names, addresses, medical records or financial data. If this data falls into the wrong hands, it could be exploited for malicious activities, such as identity theft or extortion.
- Re-identification risks: Although data is anonymized to protect privacy, there is a risk that it could be re-identified through linkage techniques or cross-analysis with other data sources. This could reveal the identity of the people behind the data, posing a threat to their privacy.
- Model theft: AI models trained on valuable data can be a target for intellectual property theft. If the models are stolen, they could be used by malicious actors for their own benefit or unfair competition.

Another problem associated with the use of AI is the decrease in job vacancies in sectors undergoing automation processes. Some key aspects of this challenge are detailed below [41, 22]:

- Job displacement: Automation through AI and advanced technologies may gradually replace certain tasks or jobs, leading to the displacement of workers in those areas. This may cause unemployment and generate economic insecurity for people who lose their jobs due to automation.
- Retraining and reskilling: AI implementation may require new skills and knowledge to operate and maintain these systems. Displaced workers will need retraining and reskilling opportunities to compete in a rapidly evolving labor market.
- Economic inequality: The adoption of AI and automation could exacerbate economic and social gaps. Those with the right skills to work with advanced technologies may benefit, while others may face difficulties finding employment.
- Impact on specific industries: Some industries and sectors may be more susceptible to automation than others. Routine and repetitive tasks in areas such as manufacturing, transportation, and customer service are more likely to be replaced by AI.
- Adapting labor policies: Governments and companies will need to develop labor and social policies to address the effects of automation on employment. This could include training and retraining initiatives, as well as measures to protect displaced workers.

Taking responsibility for decisions made by artificial intelligences (AI) is a major challenge in the field of security and ethics [56]. As AI becomes more autonomous and makes decisions that impact various aspects of society, it is crucial to determine who is responsible for the resulting actions and consequences. Some key aspects of this challenge are [64]:

- Lack of regulation and legal framework: As AI continues to advance, many jurisdictions have yet to establish clear laws and regulations on liability for AI decisions. This can lead to uncertainty and loopholes in the event of accidents or problematic situations.
- Shared responsibility: In some cases, responsibility for AI decisions may be divided among several actors, including developers, system owners, data providers, and users. Determining the specific responsibilities of each party can be complex.
- Change in the nature of error: While human errors may be more understandable and attributable, AI errors can result from complex and non-intuitive interactions between the system and the training data. This makes it difficult to assign responsibility should an error occur.
- Life-and-death decisions: In critical applications such as autonomous vehicles or healthcare systems, decisions made by AI can have direct consequences on people's lives. Determining who is responsible in these situations is particularly complex and sensitive.

To address this challenge, it is important that developers, users and regulators work together to establish clear legal and ethical frameworks that define liability in the use of AI. In addition, it is essential to encourage transparency in the development of AI algorithms and systems to facilitate understanding of their decisions and ensure that measures are taken to mitigate bias and error. Ethics and accountability in the design and deployment of AI are critical to harnessing its benefits and minimizing the associated risks.

3.1 Ethics in the development of artificial intelligence for cybersecurity

To ensure ethics in the field of cybersecurity, frequently cybersecurity specialists adhere to codes of ethics. Cybersecurity codes of ethics are documents that set out the principles and ethical standards that professionals must follow in the exercise of their profession. These codes aim to promote ethical behavior, responsibility and integrity in the handling of information and systems security. Although the content may vary according to the organization or entity issuing it, some common points that could be included are [5]:

- Confidentiality: cybersecurity specialists must protect confidential information to which they have access during the course of their work and not disclose it without proper authorization.
- Data integrity: They must ensure the accuracy and precision of the data they handle, avoiding unauthorized alterations.
- Availability of systems: Specialists must ensure that systems and services are available and function properly for authorized users.
- Professional responsibility: They must perform their work diligently and competently, acting in the best interest of the client or organization.
- Compliance with laws and regulations: They must respect and comply with the laws and regulations applicable in their field of performance.
- Conflict of interest: Avoid situations that may generate conflicts of interest or compromise their objectivity and impartial judgment.
- Do not violate privacy: Do not take actions that invade the privacy of individuals, unless strictly necessary and in accordance with applicable laws.
- Collaboration and responsible disclosure: Promote responsible sharing of vulnerability and threat information in the cybersecurity community to improve overall security.
- Respect for copyrights and intellectual property: Do not use, copy or distribute software or other copyrighted materials without proper authorization.
- Transparency: Be transparent and honest in communicating with clients and employers about the risks and limitations of security systems.

Cybersecurity professionals are generally required to review and follow codes of ethics established by their employer, professional association or appropriate regulatory body. Conveying ethical principles to an AI performing cybersecurity tasks is an important process to ensure that the AI acts in a responsible manner that respects the privacy and security of systems [38]. Here are some ways to convey those ethical principles to an AI [5]:

- Ethical design: ethical principles should be incorporated from the beginning of AI design. Engineers and developers should consider the ethical implications of design decisions, such as privacy, transparency, and respect for user rights [72].
- Ethical dataset: When training AI, a dataset that reflects ethical principles and is free of bias and discrimination should be used. This will prevent the AI from learning undesirable or discriminatory behavior [61].
- Behavioral guidelines: AI should be programmed with algorithms and guidelines that reflect ethical principles. For example, there may be explicit rules to protect data privacy and respect laws and regulations.
- Monitoring and auditing: It is important to monitor and audit AI behavior in real time to ensure that it follows established ethical principles. This involves monitoring its activity and taking action in case ethical issues arise [47].
- Continuous learning and adaptation: AI must be prepared to learn and adapt based on new ethical guidelines or changes in rules and regulations [47].
- Interaction with humans: If AI interacts with humans, it must do so in an ethical and respectful manner. This may involve setting clear limits on the type of information the AI can collect and how it can use it [47].
- Ethical testing: Before implementing an AI in production environments, it is important to subject it to ethical testing to evaluate its behavior and detect potential ethical issues [72].

- Transparency: AI systems should be designed so that their decisions and actions are understandable to humans. Transparency allows for greater accountability and makes it easier to identify bias or inappropriate behavior.
- Developer accountability: AI developers and managers must take responsibility for ensuring that AI acts in accordance with established ethical principles.
- Ethics training: Development teams and cybersecurity specialists working with AI should receive ethics training to understand the challenges and ethical implications of their actions [72].

Among the laws and regulations that exist and must be respected by both cybersecurity specialists and the IAs developed for these tasks are the following [56]:

- Data protection laws: Many countries have data protection laws that establish how personal data should be treated and protected. AI used in cybersecurity must comply with these regulations to ensure privacy and confidentiality of information [61].
- Cybersecurity laws: Some countries have specific laws to regulate the security of computer systems and networks. These laws may require certain security and safety standards for organizations and companies operating in that country [29].
- Liability laws: AI used in cybersecurity must be subject to liability and civil liability laws in the event of a security breach or incident affecting third parties [9].
- Intellectual property laws: Intellectual property regulations should apply to AI algorithms and technologies used in cybersecurity, protecting the rights of the owners of such technologies [8].
- Sector-specific regulations: Some specific sectors, such as finance or healthcare, may have additional regulations that must affect the use of AI in cybersecurity.

The elucidated laws are not exclusive to cybersecurity experts and AI systems developed for that purpose; rather, they encompass the same legal framework applied to cybercriminals. This legal paradigm extends across various countries, where penal codes delineate a spectrum of cybercrimes, ensuring that the principles and regulations designed to address illicit activities in the digital domain are uniformly applicable. In essence, these laws serve as a comprehensive framework, transcending distinctions between those safeguarding digital security and those engaging in cyber malfeasance, establishing a standardized approach to combat cyber threats on a global scale. [68]

In the cybercrime landscape, where the clandestine actions of proficient hackers often elude identification, the Budapest Convention on Cybercrime [69, 15] emerges as a pivotal response to these challenges. This international treaty, primarily tailored for European nations, outlines a comprehensive framework for addressing cyber threats and offenses. It establishes a set of legal measures and cooperation mechanisms to combat various forms of cybercrime. Notably, its impact extends beyond Europe, with studies and analyses conducted in countries like Peru [49] and Chile [7] in Latin America, reflecting the global relevance of its principles. Some of the crimes classified in these countries are the following [28, 7, 49]:

- Unauthorized access to computer systems
- Computer fraud
- Facilitation of means for computer fraud
- Computer espionage
- Unauthorized disclosure of private data or content
- Receipt of computer data
- Dissemination of programs intended to damage or interrupt
- Damage to computer or telematic systems

- Violation, theft, and deletion of correspondence
- Illegal interception, hindrance, or interruption of communications
- Falsification, alteration, or deletion of the content of computer
- Possession of child pornography

Despite the existence of well-defined cybercrime statutes and laws, there are instances where a proficient hacker adeptly avoids leaving traces, shrouding their identity in a veil of anonymity. Consequently, determining the perpetrator of an attack or identifying the individual who violated the law becomes a formidable challenge. In such cases, the absence of apparent accountability underscores the elusive nature of these skilled individuals, leaving law enforcement and cybersecurity experts grappling with the complexities of attribution in the digital landscape [77].

It is important to note that the legal and regulatory landscape around AI is constantly evolving and may vary by country and region. New regulations specific to AI in cybersecurity may have emerged after my date of knowledge. Therefore, I recommend you consult updated and legal sources for more accurate and current information on current regulations in this field [2].

3.2 Impact of ethics on artificial intelligence for cybersecurity development

Cybersecurity with AI presents a number of ethical challenges and constraints that can affect its effectiveness and application [2]. Some ways in which ethical constraints can impact AI cybersecurity techniques include the following [56, 4]:

- Constraints on data collection and use: Ethical constraints on the collection and use of personal data can limit the amount and quality of data available to train AI models. This can affect the ability of AI systems to effectively identify and mitigate threats.
- Bias and discrimination: Ethical consideration of avoiding bias and discrimination in AI systems may lead to restricting or having to adjust training with historical data. This may affect the accuracy and efficacy of models in certain contexts.
- Transparency and explainability: Requiring transparency and explainability of AI models may limit the use of more complex and difficult-to-interpret artificial intelligence techniques. Simpler but less precise approaches may be preferred to facilitate understanding and accountability.
- Security vs. privacy dilemma: In some cases, there may be an ethical dilemma between ensuring security and protecting private data. The need to maintain privacy and protect the rights of individuals may restrict certain cybersecurity practices involving analysis of or access to sensitive data.
- Limitations on experimentation and testing: Ethical restrictions may make it difficult to test in real-world environments or experiment with potential threats to evaluate the effectiveness of AI techniques. This can lead to a lower level of confidence in AI cybersecurity systems prior to implementation.

In summary, ethical constraints may affect the development and implementation of intelligent AI cybersecurity techniques, as ethical values and principles must be taken into account when making decisions about the design, training, and use of these systems. However, it is also important to note that ethical consideration is essential to ensure that AI in cybersecurity is used responsibly, fairly, and with respect for the rights of the users and individuals involved [56].

By complying with ethical constraints and sound ethical principles, trust in AI applications in cybersecurity is fostered. Trust is a crucial factor in the adoption and acceptance of these technologies. Ethical constraints establish clear responsibilities in the development and use of AI in cybersecurity. This helps ensure that organizations and individuals are accountable for their actions and decisions should incidents or problems occur. Irresponsible or negligent practices in the implementation of AI in cybersecurity are avoided and ethical issues can be identified and resolved before they become significant obstacles. This provides greater adaptability and resilience to AI-enabled cybersecurity systems.

3.2.1 Open Source Artificial Intelligence Techniques and Cybersecurity

One of the key challenges in the field of cybersecurity is the use of open-source artificial intelligence techniques. The open-source approach brings transparency and allows for public review, fostering collaboration in projects that can lead to continuous improvements in cybersecurity techniques. The involvement of multiple experts facilitates the contribution of ideas and enhancements, thereby accelerating the development of security solutions. The availability of open-source code also streamlines the implementation of these solutions, enabling organizations to adapt efficiently to new threats [35, 74].

However, the accessibility of the source code can be a double-edged sword, as it carries the risk of exposing potential weaknesses in the system to potential attackers. They can analyze open-source code to identify vulnerabilities and design specific attacks. Open-source tools are used by both defenders and attackers, providing the latter with a means to understand existing defenses and improve their tactics [35, 74].

In summary, the use of free software or open-source software in the artificial intelligence techniques for the cybersecurity offers advantages in terms of the transparency and the collaboration, but also poses challenges in terms of the exposure to threats and the dependence on the community.

3.3 Artificial Intelligence and Its Role in Hacking: An Ethical and Unethical Perspective

Malicious agents are seeking to access private information for illegal or profit-driven purposes; these are malicious hackers or cybercriminals. These Black Hat Hackers are often experts in computer systems who break into an individual's or company's devices and networks with malicious intent to steal or damage their information, compromising their security. They violate computer security for personal gain. These are individuals who often want to demonstrate their extensive knowledge of computers and commit various cybercrimes such as identity theft, credit card fraud, and denial of service attacks, among others [30].

To prevent this, experts in systems, especially in network and internet security protocols, use their knowledge to carry out tests that identify vulnerabilities and overcome an organization's security measures; these individuals are known as ethical hackers or pen-testers. The latter conduct security audits through procedures such as penetration testing, vulnerability testing, or simply "pen tests" with prior approval from the organization. The results are delivered to the system administrators, who implement improvements to eliminate risks and threats. In summary, ethical hackers are essential to ensure the integrity and reliability of computer and computing systems [31].

Artificial intelligence has a multitude of benefits and uses in cybersecurity. With the rapid evolution of cyberattacks and the proliferation of devices in today's world, artificial intelligence and machine learning can help keep up with cybercriminals, automate threat detection, and respond more efficiently than traditional software-based or human-driven methods [67].

In the realm of ethical hacking, efforts are made to exploit digital traces left by cybercriminals; this is known as intrusion signatures. Pentesters with expertise in AI build large datasets from these traces to feed algorithms that help identify attacker flaws and habits for detection and prevention. An AI system can be trained to detect intrusions in real-time if there is a sufficiently large library of these traces and patterns. Some examples of cybersecurity uses are [62, 3]:

- Identification and prevention of spam and fraudulent emails: Several email platforms like Outlook or Gmail use Artificial Intelligence to identify spam or fraudulent emails. These AIs are trained by millions of users who provide feedback, correcting classification decisions made by the algorithm in real-time. As a result, these algorithms can identify even the most subtle spam emails that attempt to go unnoticed.

- Fraud detection: Many banking entities use AI-based fraud detection systems that utilize algorithms based on expected consumer behavior to identify fraudulent transactions. These systems examine normal customer purchase patterns, the seller, transaction location, and many other data points to determine if a purchase is unusual.

- Botnet detection: Botnet detection is a highly complex area often relying on pattern recognition and proxy server synchronization analysis. Botnets are typically controlled by a master script of instructions, so a large-scale botnet attack usually involves a large number of "users" performing the same query to a server during a single attack and is very difficult to contain or even identify by a human. Hence, automated systems are used to identify these connection patterns and allow for almost instant identification of a DDoS attack and taking action accordingly.

These are just some of the areas where artificial intelligence has been utilized in cybersecurity. There is a wealth of research articles providing compelling data supporting the effectiveness of artificial intelligence in the field of cybersecurity. According to most research studies, the success rate in identifying cyberattacks ranges from 85 to 99 percent. In [40], a system of Artificial Neural Networks (ANN) with backpropagation is proposed for identifying spam SMS messages, achieving an accuracy of 95.81%. This approach is compared with other algorithms such as Support Vector Machines, Random Forests, or K-Nearest Neighbors, where accuracies range between 60% and 90% in all cases. Additionally, in [66], a comparison is conducted among various email spam classification algorithms, considering three different techniques: Abundance of Advanced Features, Enhanced Feature Selection, and an Ensemble Spam Classifier. In this analysis, a higher accuracy of 89% is achieved using the latter mentioned method.

In [24], a methodology for credit card fraud detection using machine learning techniques is introduced. A neural network based on an isolation forest, achieving an accuracy of 95%, is compared with a linear regression system, reaching an accuracy of 98%. These accuracy rates are considerably higher compared to another employed algorithm, specifically a K-Means model, which achieves an accuracy of 54%. Another example of a system for fraud detection is found in [58], where a framework for predicting fraudulent transactions using Graph Neural Networks (GNN) designed for online platforms is proposed.

Recently, there has been a trend in studying the possibilities that arise from combining classical botnet detection mechanisms with statistical machine learning techniques. These methods include the use of supervised and unsupervised algorithms to make intelligent decisions based on collected packet features, and in many cases, they have the potential to outperform traditional botnet detection methods [23]. For example, a botnet detection system based on neural networks demonstrated an accuracy of 98.6% in tests conducted in the work [18]. In another method developed to detect the spread of botnets on IoT devices, an accuracy of 97.3% was achieved using a logistic regression model [57].

But what happens when cybercriminals use artificial intelligence in their cyberattacks?

There is a significant risk that cybercriminals can launch their own AI-driven cyberattacks. The DARPA Cyber Grand Challenge, a hackathon, was one of the first to show what an AI-driven cyberattack could look like. Several teams in this competition managed to carry out automated cyberattacks, including the creation of vulnerabilities, patch production, and targeted attacks. This malicious use of artificial intelligence increases the speed and success rate and expands the capabilities of attacks. Furthermore, hackers are capable of deceiving AI-based systems in various ways. As an example, a group of researchers demonstrated that they could trick autonomous vehicles by abusing the vehicles' traffic sign recognition system. Using simple tools like graffiti and art objects, they managed to convince the cars to misinterpret traffic signs. To deceive AI-based cybersecurity, cybercriminals must first attack the categorization algorithms that AI has learned to recognize and exploit [36].

3.4 Ethics of AI Impact on Society and Public Trust

Trust is a fundamental human mechanism necessary to address vulnerability, uncertainty, complexity, and ambiguity in situations collectively constituting a risk. Within the context of trust in Artificial Intelligence, it is important to clearly define two points. First, trust in Artificial Intelligence, which refers to a computer's ability to make decisions with a certain degree of independence, and second, trust in the companies and institutions that implement and use these systems, often in opaque ways for end-users [42]. The autonomous functioning of AI has shifted the balance of power between humans and machines, making it necessary for humans to trust technology. Furthermore, the deep learning algorithms that power today's AI lack sufficient transparency and explainability, adding difficulty to the challenge of creating trustworthy technology for the public [71]. To gain trust in AI systems, certain requirements must be met [26]:

- Human Mediation and Supervision: There must be a mechanism for human supervision through a human-machine interaction approach.

- Technical Robustness and Security: AI systems must be secure, reliable, and reproducible to minimize unintended harm.

- Data Privacy and Data Governance: Data privacy and protection must be ensured, requiring an appropriate data governance framework.

- Transparency: AI systems must be transparent, and their decisions must be explainable to stakeholders. People should be informed about the capabilities and limitations of the systems.

- Diversity, Non-discrimination, and Equity: AI systems must be accessible to all, and unfair biases must be avoided.

- Social and Environmental Well-being: AI systems must benefit humans and consider the social impact and environmental consequences of their decisions.

- Responsibility: Mechanisms must be established to ensure the accountability of AI systems and their outcomes.

Similarly, trust in technology companies has become a widely debated topic today. A statistical study revealed that Americans perceive technology companies to have a more positive impact on society than other institutions, such as the media and the government. However, their feelings about the handling of ethics by technology companies have dramatically declined since 2015. There is also considerable variation in how the public views individual technology companies; a consistent finding in surveys is that the public largely distrusts Facebook/Meta, especially regarding the handling of personal data [84].

It is also essential to highlight the vast amount of personal data we feed into large algorithms every day, destined to arbitrarily create a characteristic profile of each individual, including their likes, preferences, trends, and any information that enables targeted advertising for the purchase of certain products or the focus of electoral campaigns on specific audiences. Most of these data are collected through online interactions, often hidden behind intentionally opaque "Terms and Conditions" or with mandatory cookie systems that are not easily understood by the average user, who may not be aware of data privacy and security. They often click the button just to access the information they need [60].

This massive amount of information collected by large companies is, in most cases, not used to improve the service provided to users. Typically, it is utilized for generating personalized ads, targeted marketing campaigns, and, in general, mechanisms for maximizing profits by companies or third parties to whom all this data is sold. This accumulation of personal information also presents a highly attractive target for attackers interested in accessing all this data, as with just one breach, they can gain access to the personal information of millions of people. An example of this is the sale of the data of 700 million LinkedIn users on the DarkWeb in 2021, or the leak suffered by Yahoo! in 2013 where the names, email addresses, phone numbers, dates of birth, hashed passwords, and, in some cases, security questions and their respective answers of up to 3 billion accounts were exposed [13, 79]. Although the responsibility falls almost entirely on the companies, it is important that users are aware of the information they are sharing, with whom, and for what purpose.

Society distrusts artificial intelligence systems, but it has gradually come to accept algorithms making potentially important decisions in daily life as a daily norm. Organizations developing artificial intelligence have the responsibility to implement AI systems properly and ensure they comply with ethical standards. However, to achieve public trust, an authority is needed to compel organizations to take these responsibilities seriously and validate their interpretations of these standards. Public trust requires the development of a broader infrastructure that creates conditions for trustworthy AI to thrive: a whole system of rules is needed, along with resources to train adequately qualified individuals to enforce these rules [60].

4 Conclusions

Ethics stands as a fundamental pillar in the development of Artificial Intelligence (AI), outlining key principles such as transparency, fairness, privacy, accountability and beneficence. Cybersecurity, a prominent beneficiary of AI, has improved the detection and response to threats, covering areas such as spam identification, bank fraud detection and botnet identification, although the opacity in the algorithms poses ethical challenges.

Privacy and security of data, essential in the training of AI models to prevent attacks, are threatened by risks of leaks and model theft. Transparency, real-time supervision and ethical training emerge as crucial factors for the responsible action of AI in cybersecurity.

The open source approach of AI for cybersecurity facilitates transparency and collaboration, but entails risks of exposing weaknesses to attackers. The intersection between AI and cybersecurity reveals a dual landscape, with ethical applications, such as "ethical hackers" and potential threats, such as the malicious use of AI by cybercriminals.

Although AI plays a crucial role in the detection and prevention of threats, the possibility of cybercriminals using AI poses significant risks. The decline in trust in the ethical management of technology companies is linked to the concern for the privacy of data collected in an opaque way. The responsibility lies with these companies to implement ethical and transparent systems. The massive accumulation of data presents risks both in security and in its use. The lack of awareness about privacy and security, coupled with opaque practices, creates an environment conducive to violations and exploitation of data. Compliance with laws and regulations, such as those for data protection, is imperative for an ethical deployment of AI in cybersecurity. Although ethical restrictions may affect the collection and use of data, it is necessary to weigh the biases caused by such restrictions, seeking to ensure accountability and trust.

To sustain trust in AI and its contribution to cybersecurity, a robust infrastructure with clear rules, resources for the training of professionals and the effective application of ethical standards is required. Public trust not only rests on the ethics of technological development, but on a normative framework that ensures accountability and ethical compliance by the implementing organizations of these technologies.

Acknowledgements

Acknowledgments to the Facultad de Informática de la Universidad Complutense de Madrid for allowing us to carry out a mobility stay with their researchers from the Department of Software Engineering and Artificial Intelligence, especially Dr. Juan Pavón. We also extend our thanks to the Comisión de Operación y Fomento de Actividades Académicas del Instituto Politécnico Nacional for their support in the realization of this Mobility, and to the Centro de Investigación en Computación and the Escuela Superior de Ingeniería Mecánica y Eléctrica Unidad Culhuacan.

References

[1] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion*, 99:101805, 2023.

[2] Jozef Andrasko, Matus Mesarcik, and Ondrej Hamulak. The regulatory intersections between artificial intelligence, data protection and cyber security: challenges and opportunities for the eu legal framework. *AI & SOCIETY*, 36:623–636, 2021.

[3] Paul S. Andrews and Jon Timmis. On diversity and artificial immune systems: Incorporating a diversity operator into ainet. In Bruno Apolloni, Maria Marinaro, Giuseppe Nicosia, and Roberto Tagliaferri, editors, *Neural Nets*, pages 293–306, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[4] Meraj Farheen Ansari, Bibhu Dash, Pawankumar Sharma, and Nikhitha Yathiraju. The impact and limitations of artificial intelligence in cybersecurity: A literature review. *International Journal of Advanced Research in Computer and Communication Engineering*, 11(9):127, 2022.

[5] Raj Badhwar. Ai code of ethics for cybersecurity. In *The CISO's Next Frontier: AI, Post-Quantum Cryptography and Advanced Security Paradigms*, pages 41–44. Springer, 2021.

[6] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

[7] G. Bascur and R. Peña Sepúlveda. Los delitos informáticos en chile: Tipos delictivos, sanciones y reglas procesales de la ley 21.459, primera parte. *Revista De Estudios De La Justicia*, 2023(37), 2023.

[8] Lionel Bently, Brad Sherman, Dev Gangjee, and Phillip Johnson. *Intellectual property law*. Oxford University Press, 2022.

[9] Lucas Bergkamp. *Liability and environment: private and public law aspects of civil liability for environmental harm in an international context*. Brill, 2021.

[10] Nick Bostrom and Eliezer Yudkowsky. The ethics of artificial intelligence. In *Artificial intelligence safety and security*, pages 57–69. Chapman and Hall/CRC, 2018.

[11] Kirk Bresniker, Ada Gavrilovska, James Holt, Dejan Milojicic, and Trung Tran. Grand challenge: Applying artificial intelligence and machine learning to cybersecurity. *Computer*, 52(12):45–52, 2019.

[12] Miles Brundage, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff, Bobby Filar, et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*, 2018.

[13] JesÃºs CalderÃ³n. $Nuevafiltraci\tilde{A}^3ndedatosenlinkedin, Fechadeacceso : Octubre30, 2023$.

[14] Nicola Capuano, Giuseppe Fenza, Vincenzo Loia, and Claudio Stanzione. Explainable artificial intelligence in cybersecurity: A survey. *IEEE Access*, 10:93575–93600, 2022.

[15] L.Y.C. Chang. *Legislative Frameworks Against Cybercrime: The Budapest Convention and Asia*, chapter 6, pages 327–343. Palgrave Macmillan, 2020.

[16] Chian-Hsueng Chao. Ethics issues in artificial intelligence. In *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pages 1–6, 2019.

[17] F. Charmet, H.C. Tanuwidjaja, S. Ayoubi, et al. Explainable artificial intelligence for cybersecurity: a literature survey. *Annals of Telecommunications*, 77:789–812, 2022.

[18] Shao-Chien Chen, Yi-Ruei Chen, and Wen-Guey Tzeng. Effective botnet detection through neural networks on convolutional features. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pages 372–378, 2018.

[19] Mark Coeckelbergh. Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and engineering ethics*, 26(4):2051–2068, 2020.

[20] Mark Coeckelbergh. Time machines: Artificial intelligence, process, and narrative. *Philosophy & Technology*, 34(4):1623–1638, 2021.

[21] Angelo Corallo, Mariangela Lazoi, Marianna Lezzi, and Angela Luperto. Cybersecurity awareness in the context of the industrial internet of things: A systematic literature review. *Computers in Industry*, 137:103614, 2022.

[22] Timothy Coupe. Automation, job characteristics and job insecurity. *International Journal of Manpower*, 40(7):1288–1304, 2019.

[23] Xiaxin Dong, Jianwei Hu, and Yanpeng Cui. Overview of botnet detection based on machine learning. In *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, pages 476–479, 2018.

[24] Vaishnavi Nath Dornadula and S Geetha. Credit card fraud detection using machine learning algorithms. *Procedia Computer Science*, 165:631–641, 2019. 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019.

[25] Natalia DÃaz Rodriguez, Javier Del Ser, Mark Coeckelbergh, Marcos Lopez de Prado, Enrique Herrera Viedma, and Francisco Herrera. Connecting the dots in trustworthy artificial intelligence: From ai principles, ethics, and key requirements to responsible ai systems and regulation. *Information Fusion*, 99:101896, 2023.

[26] Karen Elliott, Rob Price, Patricia Shaw, Tasos Spiliotopoulos, Magdalene Ng, Kovila Coopamootoo, and Aad van Moorsel. Towards an equitable digital society: artificial intelligence (ai) and corporate digital responsibility (cdr). *Society*, 58(3):179–188, 2021.

[27] Wolfgang Ertel. *Introduction to artificial intelligence*. Springer, 2018.

[28] Leandro Ezequiel Fusco. Los delitos informáticos en el código penal italiano. *Derecho Global. Estudios sobre Derecho y Justicia*, V(14):127–149, 2020.

[29] Gloria Gonzalez Fuster and Lina Jasmontaite. Cybersecurity regulation in the european union: the digital, the critical and fundamental rights. *The ethics of cybersecurity*, pages 97–115, 2020.

[30] Foram Gandhi, Drashti Pansaniya, and Swapna Naik. Ethical hacking: Types of hackers, cyber attacks and security. *International Research Journal of Innovations in Engineering and Technology*, 6(1):28, 2022.

[31] Aman Gupta and Abhineet Anand. Ethical hacking and hacking attacks. *Int. J. Eng. Comput. Sci*, 6(6):2319–7242, 2017.

[32] Thilo Hagendorff. The ethics of ai ethics: An evaluation of guidelines. *Minds and machines*, 30(1):99–120, 2020.

[33] Anand Handa, Ashu Sharma, and Sandeep K Shukla. Machine learning in cybersecurity: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4):e1306, 2019.

[34] Swetha Hariharan, Anusha Velicheti, A.S. Anagha, Ciza Thomas, and N. Balakrishnan. Explainable artificial intelligence in cybersecurity: A brief review. In *2021 4th International Conference on Security and Privacy (ISEA-ISAP)*, pages 1–12, 2021.

[35] Nihad A Hassan and Rami Hijazi. *Open source intelligence methods and tools*. Springer, 2018.

[36] AMA Hawamleh, Almuhannad Sulaiman M Alorfi, Jassim Ahmad Al-Gasawneh, and Ghada Al-Rawashdeh. Cyber security and ethical hacking: The importance of protecting user data. *Solid State Technology*, 63(5):7894–7899, 2020.

[37] Patrick Henz. Ethical and legal responsibility for artificial intelligence. *Discover Artificial Intelligence*, 1:1–5, 2021.

[38] John Howard. Artificial intelligence: Implications for the future of work. *American Journal of Industrial Medicine*, 62(11):917–926, November 2019.

[39] T. Irwin. *Nicomachean Ethics*. Hackett Publishing Company, Incorporated, 2019.

[40] Ankit Kumar Jain, Diksha Goel, Sanjli Agarwal, Yukta Singh, and Gaurav Bajaj. Predicting spam messages using back propagation neural network. *Wireless Personal Communications*, 110(1):403 â 422, 2020.

[41] Akanksha Jaiswal, C. Joe Arun, and Arup Varma. Rebooting employees: upskilling for artificial intelligence in multinational corporations. *The International Journal of Human Resource Management*, 33(6):1179–1208, 2022.

[42] Anna Jobin, Marcello Ienca, and Effy Vayena. The global landscape of ai ethics guidelines. *Nature machine intelligence*, 1(9):389–399, 2019.

[43] Nektaria Kaloudi and Jingyue Li. The ai-based cyber threat landscape: A survey. *ACM Comput. Surv.*, 53(1), feb 2020.

[44] Ramanpreet Kaur, Dušan Gabrijelčič, and Tomaž Klobučar. Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion*, 97:101804, 2023.

[45] C.M. Korsgaard, M. Gregor, and J. Timmermann. *Kant: Groundwork of the Metaphysics of Morals*. Cambridge Texts in the History of Philosophy. Cambridge University Press, 2012.

[46] Alekya Sai Laxmi Kowta, P. K. Harida, Shruti Varsha Venkatraman, Siddharth Das, and V. Priya. Cyber security and the internet of things: Vulnerabilities, threats, intruders, and attacks. In Nabendu Chaki, Nagaraju Devarakonda, Agostino Cortesi, and Hari Seetha, editors, *Proceedings of International Conference on Computational Intelligence and Data Engineering*, pages 387–401, Singapore, 2022. Springer Nature Singapore.

[47] Joshua A. Kroll, James Bret Michael, and David B. Thaw. Enhancing cybersecurity via artificial intelligence: Risks, rewards, and frameworks. *Computer*, 54(6):64–71, 2021.

[48] Pawan Kumar and Manjit Singh. Ethical challenges of using artificial intelligence in cybersecurity. In Pawan Kumar and Manjit Singh, editors, *Cyber Laws in India: Emerging Trends*, pages 403–411. University Book House Pvt. Ltd., Jaipur, Forthcoming.

[49] C. Leyva Serrano. Estudio de los delitos informáticos y la problemática de su tipificación en el marco de los convenios internacionales. *Lucerna Iuris Et Investigatio*, 2021(1):29–47, 2021.

[50] S Matthew Liao. *Ethics of artificial intelligence*. Oxford University Press, 2020.

[51] Caroline Lloyd and Jonathan Payne. Rethinking country effects: robotics, ai and work futures in norway and the uk. *New Technology, Work and Employment*, 34(3):208–225, November 2019.

[52] Katanosh Morovat and Brajendra Panda. A survey of artificial intelligence in cybersecurity. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 109–115, 2020.

[53] Vincent C Müller. Ethics of artificial intelligence and robotics. *Stanford Encyclopedia of Philosophy*, 2020:1–31, 2020.

[54] Thanh Thi Nguyen and Vijay Janapa Reddi. Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–17, 2021.

[55] Nils J Nilsson. *Principles of artificial intelligence*. Springer Science & Business Media, 1982.

[56] Shane O'Sullivan, Nathalie Nevejans, Colin Allen, Andrew Blyth, Simon Leonard, Ugo Pagallo, Katharina Holzinger, Andreas Holzinger, Mohammed Imran Sajid, and Hutan Ashrafian. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (ai) and autonomous robotic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 14:e1968, 2018.

[57] Anton O. Prokofiev, Yulia S. Smirnova, and Vasiliy A. Surov. A method to detect internet of things botnets. In *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pages 105–108, 2018.

[58] Susie Xi Rao, Shuai Zhang, Zhichao Han, Zitao Zhang, Wei Min, Zhiyao Chen, Yinan Shan, Yang Zhao, and Ce Zhang. Xfraud: Explainable fraud transaction detection. *Proc. VLDB Endow.*, 15(3):427â436, nov 2021.

[59] C.D.C. Reeve. *Republic*. Hackett Classics. Hackett Publishing Company, Incorporated, 2004.

[60] Pedro Robles and Daniel J Mallinson. Artificial intelligence technology, public trust, and effective governance. *Review of Policy Research*, 2023.

[61] Cedric Ryngaert and Mistale Taylor. The gdpr as global data protection regulation? *American Journal of International Law*, 114:5–9, 2020.

[62] BS Sagar, S Niranjan, Nithin Kashyap, and DN Sachin. Providing cyber security using artificial intelligence–a survey. In *2019 3rd international conference on computing methodologies and communication (ICCMC)*, pages 717–720. IEEE, 2019.

[63] Sagar Samtani, Murat Kantarcioglu, and Hsinchun Chen. Trailblazing the artificial intelligence for cybersecurity discipline: A multi-disciplinary research roadmap. *ACM Trans. Manage. Inf. Syst.*, 11(4), dec 2020.

[64] Filippo Santoni de Sio and Giulia Mecacci. Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, 34:1057–1084, 2021.

[65] J.P. Sartre, C. Macomber, A. Cohen-Solal, and A. Elkaïm-Sartre. *Existentialism is a Humanism*. Yale University Press, 2007.

[66] Nasir Fareed Shah and Pramod Kumar. A comparative analysis of various spam classifications. In Pankaj Kumar Sa, Manmath Narayan Sahoo, M. Murugappan, Yulei Wu, and Banshidhar Majhi, editors, *Progress in Intelligent Computing Techniques: Theory, Practice, and Applications*, pages 265–271, Singapore, 2018. Springer Singapore.

[67] Arab Mohammed Shamiulla. Role of artificial intelligence in cyber security. *International Journal of Innovative Technology and Exploring Engineering*, 9(1):4628–4630, 2019.

[68] Gomgom TP Siregar and Sarman Sinaga. The law globalization in cybercrime prevention. *International Journal of Law Reconstruction*, 5(2):211, 2021.

[69] F. Spiezia. International cooperation and protection of victims in cyberspace: welcoming protocol ii to the budapest convention on cybercrime. *ERA Forum*, 23:101–108, 2022.

[70] Gautam Srivastava, Rutvij H Jhaveri, Sweta Bhattacharya, Sharnil Pandya, Rajeswari, Praveen Kumar Reddy Maddikunta, Gokul Yenduri, Jon G. Hall, Mamoun Alazab, and Thippa Reddy Gadekallu. Xai for cybersecurity: State of the art, challenges, open issues and future directions, 2022.

[71] Margit Sutrop. Should we trust artificial intelligence? *Trames*, 23(4):499–522, 2019.

[72] Mariarosaria Taddeo. Three ethical challenges of applications of artificial intelligence in cybersecurity. *Minds & Machines*, 29:187–191, 2019.

[73] Mariarosaria Taddeo, Tom McCutcheon, and Luciano Floridi. Trusting artificial intelligence in cybersecurity is a double edged sword. *Nature Machine Intelligence*, 1:557–560, 2019.

[74] Mariarosaria Taddeo, Tom McCutcheon, and Luciano Floridi. *Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword*, pages 289–297. Springer International Publishing, Cham, 2021.

[75] Paul Timmers. Ethics of ai and cybersecurity when sovereignty is at stake. *Minds & Machines*, 29:635–645, 2019.

[76] Trung Cao Truong, Ivan Zelinka, Jakub Plucar, Milan Candik, and Vaclav Sulc. *Artificial Intelligence and Cybersecurity: Past, Presence, and Future*, volume 1056 of *Advances in Intelligent Systems and Computing*, chapter 30, page 781. Springer, Singapore, 2020.

[77] Nicholas Tsagourias and Michael Farrell. Cyber attribution: Technical and legal approaches and challenges. *European Journal of International Law*, 31(3):941–967, August 2020.

[78] C UNESCO. Recommendation on the ethics of artificial intelligence, 2022.

[79] WeLiveSecurity. 5 filtraciones de datos en los Ãºltimos 10 aÃ±os, $Fechadeacceso : Octubre 30, 2023$.

[80] Isaac Wiafe, Felix Nti Koranteng, Emmanuel Nyarko Obeng, Nana Assyne, Abigail Wiafe, and Stephen R. Gulliver. Artificial intelligence for cybersecurity: A systematic mapping of literature. *IEEE Access*, 8:146598–146612, 2020.

[81] Victoria Wylde, Nidhi Rawindaran, Jane Lawrence, et al. Cybersecurity, data privacy and blockchain: A review. *SN Computer Science*, 3:127, 2022.

[82] Aimee van Wynsberghe. Artificial intelligence: From ethics to policy. *Panel for the Future of Science and Technology*, 2020.

[83] Muhammad Mudassar Yamin, Mohib Ullah, Habib Ullah, and Basel Katt. Weaponized ai for cyber attacks. *Journal of Information Security and Applications*, 57:102722, 2021.

[84] Baobao Zhang and Allan Dafoe. Us public opinion on the governance of artificial intelligence. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 187–193, 2020.

[85] Zhimin Zhang, Huansheng Ning, Feifei Shi, Fadi Farha, Yang Xu, Jiabo Xu, Fan Zhang, and Kim-Kwang Raymond Choo. Artificial intelligence in cybersecurity: research advances, challenges, and opportunities. *Artificial Intelligence Review*, 55:1029–1053, 2022.