MBNA Publishing House Constanta 2022

Proceedings of the International Scientific Conference SEA-CONF

SEA-CONF PAPER • OPEN ACCESS

Cognitive Security Ethics: A Utilitarian Perspective on the Psychological Profiling of Employees

To cite this article: S. Eftimie, R. Moinescu, D. Glăvan and C. Răcuciu, Proceedings of the International Scientific Conference SEA-CONF 2022, pg. 118-123.

Available online at www.anmb.ro

ISSN: 2457-144X; ISSN-L: 2457-144X

doi: 10.21279/2457-144X-22-015

Cognitive Security Ethics: A Utilitarian Perspective on the Psychological Profiling of Employees

S Eftimie[1], R Moinescu[1], D. Glăvan[1] and C Răcuciu[2]

[1] Ph. D. Student, Military Technical Academy
[2] Prof. Eng. Ph. D., Military Technical Academy

Abstract. In the context of current cybersecurity-applied artificial intelligence advancements, businesses operating in this field need to address the subject of ethics. State-of-the-art security services use emails and chats to collect valuable insights about employees. In this paper, we review the current ethical landscape surrounding this vital subject. We put forward a utilitarian perspective on privacy issues raised during the psychological profiling process. Additionally, we describe how a utilitarian artificial intelligence concept system can be used ethically in cognitive security endeavors, improving a company's security posture while maintaining individual privacy.

1. Introduction

Cognitive Security represents the utilization of artificial intelligence modeled on human reasoning processes to detect security vulnerabilities and threats [1]. Today's security analysts must focus on the organization's cybersecurity situation through cognitive tools, enhancing the rules to detect and respond to security incidents. An emerging area in this field is the usage of psychological profiling to determine human weaknesses and behavior patterns that imply unintended or intended actions that are dangerous to an entity.

Cognitive security applications also include non-technical methods to make individuals less vulnerable to manipulation and technical solutions designed to detect misinformation and misleading data [1].

Businesses use recent innovations in artificial intelligence solutions to monitor employees. Although these methods rely on surveillance during working hours, there is a thin line between personal and corporate data and a real danger of invading personal lives. In [2] and [3], we can find several examples of surveillance inside business environments. The barrier between work and personal life has lessened due to the BYOD (Bring-Your-Own-Device) policies and the new work-from-home paradigm.

Artificial intelligence algorithms create the possibility for cognitive systems to regularly mine data for meaningful information and obtain knowledge by using high-level analytics. By steadily improving practices and methods, these automatic systems learn to anticipate threats and vulnerabilities and generate more proactive solutions.

The ability to process and examine vast volumes of data means that cognitive security systems can recognize connections among data points and trends that would be impossible for humans to discover. Similar to other cognitive computing applications, self-learning security systems use technologies such as pattern identification and natural language processing to mimic the decision processes that happen inside the human brain. Such automated security systems are designed to solve problems without requiring human resources. Cognitive Security may be beneficial to prevent cyberattacks that

manipulate human attention. Such attacks sometimes referred to as cognitive hacking, are intended to affect people's actions to serve the attacker's malicious intent.

Despite all these advantages, we need to consider that AI tools are not error-free [4] [5], and their judgments should not be used to give definitive verdicts. Preferably companies should use them to improve the security posture. There is no consistent way in which the multitude of tools on the market act unitarily since each provider focuses on its competitive advantages.

Insiders are individuals who can be regarded as threats or vulnerabilities through their intended or unintended actions inside an organization. Current methods for discovering and managing such threats and vulnerabilities are ineffective [6].

Employees are likely to view monitoring as both an unreasonable intrusion on confidentiality and an inconvenience in establishing employment relationships. As a result, monitoring can affect employee productivity. Monitoring may also accentuate the perception that employees' critical feedback is not welcome, i.e., such tracking may foster a negative relationship between employer and employees. As a result, internal procedures, isolated from employee criticism, become stagnant, impacting the company's competitiveness in the market.

Reducing the risks associated with internal threats is part of the business strategy and essentially reduces the possible attack costs. Given the costs associated with implementing and maintaining a monitoring program, artificial intelligence solutions' advantages are noticeable since they bring scale economies to the whole process and reduce the human workforce associated with security responses.

Companies need to make sure that individuals do not suffer repercussions from using this technology and, at the same time, avoid legal implications as described in [7]. In [8], we can find more arguments supporting the idea that the ethical part should not be separated from cybersecurity. Ethical and legal advancements are needed to clarify these aspects in our society both in research [9] and in practice since we have current examples such as the surveillance programs in China [10] where technology has clearly been used in ways that would be viewed as privacy breaching in Western countries. Companies should identify the main values applicable to cybersecurity and use them when developing programs that enhance it.

Two questions are relevant in the process of assessing automatic psychological profiling: Is it ethical to use artificial intelligence tools to discover personal aspects about individuals? And, is it ethical that these individuals should suffer the consequences of an AI algorithm's outputs? In the following section, we will be addressing these questions in the context of three ethical frameworks for the cybersecurity field and comment on their usage feasibility on cognitive security endeavors. In section III, we will forward a utilitarian perspective on the automatic surveillance process and discuss ways in which ethical values can be maintained while improving the security posture of a company. Conclusions will follow in the last section of the paper.

2. Ethical Frameworks in Cybersecurity

Cybersecurity is a term that outlines a set of technologies and strategies to defend infrastructures. There are three main categories of technology used in cybersecurity:

- Technologies that recognize and respond to vulnerabilities and threats.
- Technologies that guarantee the authentication of the actors participating in an exchange of information and the confidentiality of that information.
- Technologies that identify and respond to violations perpetrated in cyberspace, such as deception or falsification.

Various ethical issues arise for each type of technology. We have identified three ethical frameworks adapted for the cybersecurity area, Principlism, Human Rights and Contextual Integrity, and each one provides a broader environment in which ethical dilemmas can be resolved:

2.1. Principlism

Principlism represents a system of ethics based on a short number of principles with a grounding in common-sense morality and professional ethical practice. The principlist strategy is a simple, minimalist

framework that provides remarkable adaptability [11]. It leaves the cybersecurity agents the complex job of recognizing the particular circumstances that would resolve ethical dilemmas [11]. In [12], a cybersecurity principlist framework is described, which employs the following ethical principles:

- The principle of Beneficence helps an individual decide what is good and right in performing an action. This preference to do the right thing creates an ethical perspective that provides acceptable solutions to ethical dilemmas. This is also associated with the utility principle, which affirms that we should strive to create the most significant good instead of evil in the world [13]. The beneficence principle is primarily associated with the utilitarian ethical theory. Comparable to beneficence, the least harm principle was developed to handle circumstances in which no option seems advantageous. In events such as these, individuals attempt to choose to do the most limited potential harm and to do it to the shortest number of individuals.
- The principle of Respect for Autonomy affirms that the process of making a decision should empower individuals to be independent in making judgments that impact their own lives [12].
- The principle of Justice states that individuals should concentrate on procedures that are fair to those affected [13].

The benefit postulate applies in all generality to cybersecurity research. We argue that it would benefit our society greatly if personal aspects of individuals from an organization, let us say, were analyzed and used to create a better threat/vulnerability model for usages outside of the respective organization.

Whether individuals should suffer the consequences of the output of an AI remains an issue. We argue that current tools do not yet have the needed accuracy to make a definitive judgment, and such a tool would likely break the Justice principle. The Respect for Autonomy principle could be fulfilled though, if individuals would agree voluntarily to participate in the program.

2.2. Human Rights

The human right to privacy mandates the protection of private information [14]. In consequence, cybersecurity technology that intends to defend privacy and confidentiality is generally aligned with human rights [11]. Despite this, cases might exist in which security systems designed to protect privacy and confidentiality represent privacy threats. Cybersecurity technologies such as those used in authentication, require certification and credentials management. Credentials require the acquisition of private data concerning individuals, which may endanger users to privacy violation. In addition, other classes of cybersecurity systems, such as the ones associated with monitoring web traffic and combating cyberattacks, are in more immediate opposition to human rights [11].

To answer the question of using artificial intelligence algorithms to discover personal aspects about individuals, we will have to analyze the ethical implications of monitoring and profiling.

Surveillance presents censorship threats that can infringe the human right to free speech and eavesdropping (which can violate the human right to due process). Furthermore, surveillance is associated with profiling. In the case of discovering threats, using AI psychological profiling can be correlated with potential infringements of human rights. In profiling, people are assessed based on a set of characteristics that they possess. Although private data is used to create profiles, privacy is not the main issue. The risk is that the whole process might cause problems such as discrimination against the people involved. In consequence, in this ethical frame of reference, discovering personal aspects about individuals and use them to classify individuals inevitably leads to human rights infringement.

2.3. Contextual integrity

Contextual integrity represents an ethical framework that considers privacy from both a descriptive point of view and a normative one [11]. This theory's central tenet is that privacy infringements consist of violations of social norms concerning information transmission among individuals. The appropriate social standards are different based on the context in which the information is transmitted.

When expectations involving the way information should be communicated are respected, people’s privacy is maintained.

Although Contextual Integrity can be considered a somewhat conservative theory, the cybersecurity context may justify the introduction of new technologies such as AI, thus allowing innovation.

In [15], Nissenbaum illustrates how to use the theory as a foundation for the practical analysis of technologies perceived as privacy infringements.

The subjective doubtful feelings towards a technology (that is perceived as being problematic) are described as an outcome of its violation of expectations concerning the exchange of information in a specific social context. Consequently, the ethical justification for a surveillance process driven by artificial intelligence in the contextual integrity framework would depend upon how that information (resulting profiles) is revealed. Our society's current social norms do not allow for such information to be disclosed.

3. A Utilitarian Perspective on Psychological Profiling

Utilitarianism represents a family of normative ethical systems that guide actions that maximize satisfaction and well-being for all concerned individuals.

There is a notable cultural shift in the cybersecurity industry that recognizes the importance of individuals at all levels in an organization to participate in the security effort. If the security procedures and methods are to stop breaches successfully, they need to be adopted and exercised daily by every employee. The implication is that an organization is exposed to significant harm unless proper procedures produced at a strategic level are capable of forming the future cultural foundation of the respective entity.

The utilitarian set of ethical principles is constructed on the ability of an individual to predict the outcomes of an action. Utilitarianism states that the option that produces the most significant benefit to most individuals is the ethically correct one. The expertise used in the attempt to anticipate outcomes may be inaccurate. A utilitarian might take decisions that may seem unethical in the future if the chosen option did not serve the most people as foreseen. Two kinds of utilitarianism exist, rule utilitarianism and act utilitarianism:

Rule utilitarianism assumes respect for the law, and fairness is regarded as a central value. This means that it pursues the most significant benefit to most people but through the most honest and most law-abiding means available. The added advantages are that it incorporates beneficence and at the same time considers justice.

Act utilitarianism supports exactly the main definition of utilitarianism: performing the acts that benefit the most people, regardless of private beliefs or societal restraints such as laws.

We forward the idea that rule-based utilitarianism is an appropriate way to handle cognitive security endeavors, such as the automatic psychological profiling of employees if the system does not expose personal data, i.e., the artificial intelligence system should be designed in such a way that it can operate without human intervention, and its output should not be made available if it allows the identification of the analyzed individuals. The system would illustrate to subjects their own inadequacies so that they can be aware of them. In Fig 1. we described the characteristics of such a system. In [6], an intent-based taxonomy of insider threats and vulnerabilities was presented. Unintentional insiders could benefit from such a tool that illustrates their inadequacies. This process could help create a risk-aware culture where workers are instructed about the cybersecurity risks and be prepared to decide the right actions to shield against them. On the other hand, intentional insiders could be discouraged seeing that the system is correctly assessing their malicious activity.

This approach is in line with [16], where a human-centric approach related to cybersecurity is described. It can also be interpreted as a defence solution in line with the recommendations provided in [17].

- Such a system would have to have the following characteristics:
- An individual (and only him) can see the results of their profile assessment.
- An individual cannot be linked to his profile.
- A general statistic can be outputted from the algorithm if it does not permit the identification of any individuals and their profile.

- An independent assessment system for the security posture should be used in conjunction with the profiling system.

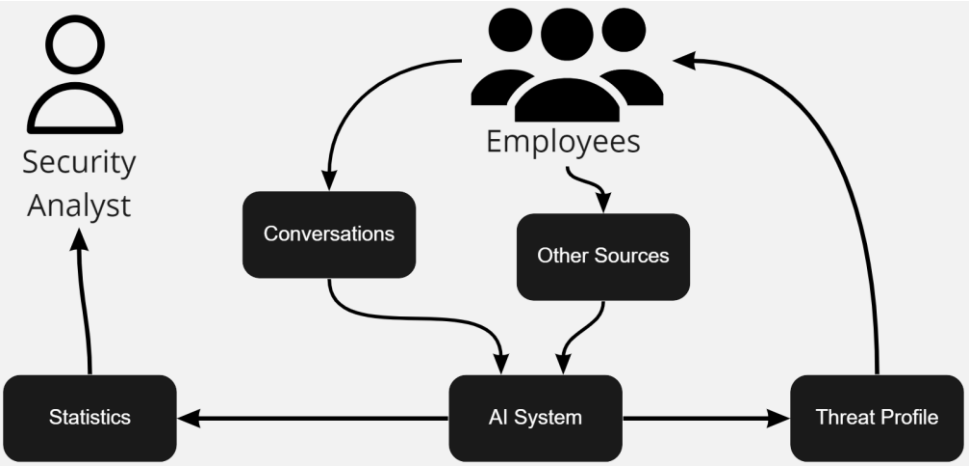

Fig. 1. Utilitarian AI Cognitive System

4. Conclusions

In this paper, we presented a set of ethical frameworks applied in cybersecurity with observations regarding their usage in the context of the latest developments in artificial intelligence-based surveillance and profiling.

We examined three non-utilitarian ethical frameworks, the principlist system that is mainly used in research endeavors, the human rights framework used in legislative efforts, and contextual integrity, which considers social norms in the assessment process.

The ethics landscape in the context of cybersecurity is a complex one. On the one hand, society must decide its goals and ethical priorities to determine what rights it wants to enact. On the other hand, systems have to be adapted in the context of the latest technological developments.

In our evaluation, a utilitarian framework is a more suitable way to address the ethical challenges associated with automatic surveillance. For this purpose, an AI-based surveillance and profiling process can be used to increase awareness regarding security while maintaining a high privacy level. The monitoring process results should be made available only to the affected subjects, enabling them to improve their cybersecurity skills.

Inadvertently, in the pursuit of security, privacy issues will emerge. We believe that a compromise in which sensitive data is left to be handled by artificial intelligence systems rather than human workers is an adequate strategy (both economically and ethically) in today's resource wars in cybersecurity.

References

[1] R. Andrade, and Y. Sang, “Cognitive security: A comprehensive study of cognitive science in cybersecurity. ” Journal of Information Security and Applications 48 (2019): 102352.

[2] M. Lafond, P. Brosseau, and Esma Aïmeur. “Privacy invasion in business environments.” 2012 Tenth Annual International Conference on Privacy, Security and Trust. IEEE, 2012.

[3] F. Thomsen, “The concepts of surveillance and sousveillance: A critical analysis.” Social Science Information 58.4 (2019): 701-713.

[4] M. Taddeo, T. McCutcheon, and L. Floridi. “Trusting artificial intelligence in cybersecurity is a double-edged sword.” Nature Machine Intelligence 1.12 (2019): 557-560.

[5] R. Yampolskiy, “Artificial intelligence safety and cybersecurity: A timeline of AI failures. ” arXiv preprint arXiv:1610.07997 (2016).

[6] S. Eftimie, R. Moinescu, and C. Răcuciu, “Insider Threat Detection Using Natural Language Processing and Personality Profiles.” 2020 13th International Conference on Communications

(COMM). IEEE, 2020
[7] P. Morrow, “The New Age Of Cybersecurity Privacy, Criminal Procedure And Cyber Corporate Ethics.” Journal of Cybersecurity Research (JCR) 3.1 (2018): 19-28.
[8] D. Shoemaker, A. Kohnke, and G. Laidlaw. “Ethics and cybersecurity are not mutually exclusive.” EDPACS 60.1 (2019): 1-10.
[9] B. Davis, C. Whitfield, and M. Anwar. “Ethical and Privacy Considerations in Cybersecurity.” 2018 16th Annual Conference on Privacy, Security and Trust (PST). IEEE, 2018.
[10] P. Timmers, “Ethics of AI and cybersecurity when sovereignty is at stake.” Minds and Machines 29.4 (2019): 635-645.
[11] C.Markus, B. Gordijn, and Michele Loi, “The Ethics of Cybersecurity”, Springer Nature, 2020, pp. 73–93
[12] D. Dittrich, and E. Kenneally, “The menlo report: Ethical principles guiding information and communication technology research.”, US Department of Homeland Security, 2012.
[13] K. Macnish, and J. van der Ham. “Ethics in cybersecurity research and practice. ” Technology in Society 63 (2020): 101382.
[14] United Nations, “The Universal Declaration of Human Rights”
[15] H. Nissenbaum, “Privacy in context: technology, policy, and the integrity of social life.”, 2009, Stanford University Press, Stanford
[16] R. Deibert, Ronald “Toward a human-centric approach to cybersecurity.” Ethics & International Affairs 32.4 (2018): 411-424.
[17] J. Pattison, “From defence to offence: The ethics of private cybersecurity.” European Journal of International Security 5.2 (2020): 233-254.