The Ethics of Social Engineering in Penetration Testing: Expanding Virtue Ethics to AI and Automated Social Engineerings

Parit Arvindbhai Jogani
Z23751881
CIS 6370-001: Computer Data Security
Professor: Eric Ackerman

December 8, 2024

Abstract

This paper examines the ethical challenges of using social engineering in penetration-testing, drawing on the virtue ethics analysis presented by [4]Joseph M. Hatfield in Virtuous Human Hacking: The Ethics of Social Engineering in Penetration-Testing. Hatfield argues that social engineering, though manipulative, can be ethically justified if aligned with virtue ethics principles—specifically the communal good and moderation. Building on this, the paper introduces a new dimension by exploring the implications of emerging technologies, particularly artificial intelligence (AI) and automated social engineering (ASE). These technologies pose unique ethical concerns, as they allow for large-scale manipulation with minimal direct human involvement, raising questions about accountability and harm. By extending the virtue ethics framework to include AI-driven practices, this paper argues that a balance must be struck between leveraging technological advancements and maintaining ethical oversight. Recommendations are offered for updating ethical guidelines in penetration-testing to address these challenges, ensuring that security practices evolve alongside technological innovations while minimizing harm.

1 Introduction

It has emerged as a major concern throughout the globe since organizational data is now an attractive target for unauthorized personnel. As a consequence of these threats, the process of penetration testing or ethical hacking has been adopted as one of the main exercises used to detect and secure likely risks before they are exploited. Penetration testing means the use of tools and techniques to test readiness of the system to resist a real attack, and among popular techniques, social engineering (using people's tricks to deceive them) is considered to be the most radical one. Social engineering is different from the technical, which involves attacking a person's cognitive weakness such as trust, fear or confusion to gain unlawful access to information. This practice can bring sever ethical dilemma due to the practice of deception and manipulating vulnerabilities found in the organizations systems, thus one has to establish the level to which ethical hackers can go while practicing Ethical hacking while still upholding their ethical standards. The evaluation of various structures of ethical theory applied to social engineering has been conducted in the previous literature primarily focusing on utilitarian and deontological theories. Different forms of utilitarianism would justify social engineering for the reason that the latter would bring in more utility compared to the costs incurred to stakeholders[7] (Finn & Jakobsson, 2007). In contrast, deontological ethics frowns at any form of distortion and was quick to assuage that it is a violation of a person's right and dignity whether the outcomes will be positive or negative. However, these approaches have been criticized for either permitting too much harm in keeping with the utilitarianism or for being inflexible to admit of usable security measures as in deontology. Abandoning the deontological view completely, [4](Joseph M. Hatfield) come up with a virtue one and opens very interesting perspective claiming that social engineering might be ethically justified if it follows the principles of such virtues as moderation and responsibility. His analysis situates ethical action within a communal frame work in which the interest of the whole community encompassing the organisation and all the employees trump the individual consent in some circumstances. However, there are areas that still need to be explored in more detail to provide satisfactory solutions when it comes to ethical considerations – the topic that Hatfield's work helps to elaborate on, yet ASE powered by AI and automation is the area that needs improvement.

In the light of advancement in technology particularly artificial intelligence, social engineering practices have adopted techniques which creates an immense environment to manipulate the targeted subjects on large scale but without physical control over the targets. This creates new ethical issues which have not been adequately realized by older structures. Ethical frameworks, such as utilitarianism, which aim to maximize the greater good while minimizing harm, provide a foundational lens for such evaluations [10](Bentham & Mill, 2009). AI-based social engineering attacks can make an attempt on one or many targets, raising questions about the scale of damage, lack of human supervision, and total non-personal approach towards the targets. While in traditional Social Engineering skills, there are agendas that contain human element and can choose how and when to proceed, AI lacks this parameter of flexibility. This paper aims to classify the considerations that are particular to AI-based penetration testing by applying [4]Hatfield's virtue ethics approach while arguing that the guidelines for the ethical use of IT have to expand to reflect the increased vulnerability caused by automation. This paper's main purpose is to implement the virtue ethics to the contemporary penetration testing with emphasis on artificial intelligence social engineering. The central research question is: In what way does the virtue ethical approach that centers around temperance and the common good help to respond to ethical problems of automated

social engineering? This paper considers social engineering within the context of penetration testing and will therefore explore both classical human-based methods and new AI based approaches. This paper will discuss the concerns between cybersecurity innovation and ethics while providing insights to penetration testers, organizations, and AI.

In briefly outlining the organization of this paper, the following sections have been developed: First, the literature review consists of present ethical theories used in social engineering and the advantages and limitations of various theories. In the case of this article, the methodology used in the evaluation of the traditional social engineering and the artificial intelligence social engineering is presented in this section. The new findings section incorporates AI distinctiveness in virtue ethical considerations and offers cases to illustrate the problems and possible remedies. Lastly, the conclusion suggests the best way through which penetration testers and organizations can use social engineering methods ethically and reduce or avoid causing harm. This paper's main argument has been that virtue ethics, whose values are moderation and the common good, still provides a solid ground to analyze the ethicality of social penetration testing. But as AI further ingrains into the cybersecurity shift, this framework must be expanded to address the given mode of automation with social engineering. Hence, starting with the outlined three key findings, this paper presents a novel framework for integrating AI ethics into virtue ethics that can guide the relationships between technology and ethics.

2 Literature Review

2.1 Introduction to Social Engineering and Ethical Challenges

Cybercrime – socially engineered penetration testing has become a critical technique during the penetration testing because it involves exploiting human weaknesses as opposed to technological defects. When described as the process of influencing a user to divulge information that would normally remain secure, social engineering differs from other hacking techniques that rely on tricking a system's security measures through pre-programmed responses such as trusting, fearing, and being bewildered[6](Mouton et al., 2016). While it is quite useful in identifying human vulnerability in security, social engineering raises questions of ethics such as privacy and consent as well as possible negative psychological effect on the participants. This ethical tension blurs the line between acceptable use of testing and ethical wrong doings in the field of security.Authors in cybersecurity studies have used utilitarianism, deontological, and virtue ethic to measure the ethicality of social engineering. Increased cases of cybercrimes demand an ethical standard that will mediate between the requirements of security and personal freedoms essential for penetration testers to work under while maintaining the sanctity of organizations.

2.2 Ethical Frameworks in Social Engineering

2.2.1 Utilitarian Ethics

The ethics theory central to cybersecurity, the utilitarianism trying to provide as much utility as possible while exemplifying as much disutility as it can [10](Bentham& Mill, 2009). In penetration testing, therefore, utilitarian ethics defends social engineering activities since their benefits outweigh the temporary harm or deceitfulness to individuals embraced by the technique [7](Finn

& Jakobsson, 2007). The first and foremost reason is that the psychological effect on people is minimal as compared to the fact that it will reduce the chances of actual attacks. But once again, you would find that utilitarian ethics has issues regarding individual autonomy or consent; what Thread with his Idiot argument did was to contend that personal harm as a result of stricter security is acceptable as a means to an end since security of many as a general populace is desirable. Distress, or mistrust in a tested organization, generated through penetration tests that engage social engineering deems utilitarianism's approach for explaining merit for harm based on the outcome as ethically questionable. Hence it is important to develop an accurate framework focusing on human rights and ethical standards within testing.

2.2.2 Deontological Ethics

Deontological ethics, based on the theory of the great philosopher Immanuel Kant, suggests that the goal is to follow the rules and principles that dictate right and wrong in any given situation [8](gregor, 1997). Deontological ethics believe like the action of deception is unethical despite its aim to serve secuity purpose, so violating individual's autonomy is wrong. According to [10](Bentham& Mill, 2009), social engineering tactics are ethically wrong if they violate an individual's information self-determination principle.

Although Global II is a systematic approach to the defense of the rights of individuals, the deontological focus on non-Deceptive practices puts constraints on the effectiveness of penetration testing. The role of actual deception could be eliminated if all test procedures strictly followed rules against deception, as many human susceptibilities are revealed only through directed manipulation. The above researchers make a projection that apply of deontological ethics is rigid, which sorts a challenge when realized in cybersecurity practices, hence the focus on virtue ethics as a more elastic theory [6](Mouton et al., 2016).

2.2.3 Virtue Ethics in Cybersecurity

While legalism and obligation-based theories mostly offer an action-based ethical frame work to which some scholars have argued that the virtue ethics character based ethical model, moderation, and responsibility are more suitable for ethical social engineering. For Aristotle, virtue ethics is based on the "golden mean" idea which enshrines a golden middle for ethical reasoning avoiding lacunae and deficiencies in people's behavor [9](Aristotle, 350 B.C.E). In the field of cybersecurity, virtue ethics provides that social engineering is good/ethical if it is done in accordance with virtues of responsibility, temperance, and commune bonum [4](Hatfield, 2020). In his work, Hatfield has incorporated virtue ethical principle of penetration testing by supporting any action that will have better outcomes for the organization members and minimal encroachment of the other people's rights.

Of all the approaches to ethical theory, virtue ethics is the most adaptable approach, which will cut across the middle between utilitarian and deontological theories to solve ethical dilemmas. In regarding the testing methods as slightly more flexible, virtue ethics helps penetration testers to perform socially valuable tests while remaining ethical. Nevertheless, this framework is not without a vice especially as AI- enhanced methods widen the frontiers of social engineering, thus, creating new forms of ethical dilemmas.

2.3 AI and Automated Social Engineering (ASE)

The integration of artificial intelligence (AI) into cybersecurity has enabled automated social engineering (ASE), which compounds ethical issues through large-scale, depersonalized manipulation. AI-driven social engineering techniques can execute massive-scale manipulations with minimal human oversight, amplifying concerns about privacy, accountability, and psychological impact on a larger scale than traditional methods . Without human discretion, AI can indiscriminately apply manipulative tactics, increasing the potential for harm while bypassing individual context [5](Floridi & Taddeo, 2016).

AI's role in social engineering complicates the application of traditional ethical frameworks. Utilitarian and deontological ethics, which both assume human decision-making, may fail to address accountability and harm at the automated scale of AI. Researchers have suggested that virtue ethics may provide a valuable foundation for developing guidelines for ASE, as it emphasizes moderation and the communal good. However, virtue ethics alone is insufficient to address challenges like impersonal harm and difficulty assigning accountability in automated contexts [3](Coeckelbergh, 2021).

2.4 Gaps in Current Literature

Although,there are prospects about applying virtue ethics as the leading paradigm for ethical collection of information through penetration testing, the application of this theoretical approach to AI-powered social engineering is uninvestigated. CS called into question by AI-based scalability and autonomy include depersonalized harm, judgemental decisionlessness, and accountability. This paper seeks to fill this gap by developing a virtue ethics model for addressing aspects of AI ethics and applying it to ASE use in cybersecurity. That is why this work aims to extend virtue ethics to incorporate automated manipulation and present a detailed approach that accounts for new ethical challenges in penetration testing.

3 Methodology

3.1 Research Design and Approach

Using a qualitative, comparative analysis approach, this study explores the ethical implications of traditional human driven social engineering (SED) and automated social engineering (ASE) in penetration testing in the context of AI driven automated social engineering (ASE). Aligning with the aim of this study is to examine how ethical principles such as virtue ethics could direct the appraisal of social engineering practices inherently manipulative but aimed at improving organizational security. Using a qualitative lens, however, the study provides means of a nuanced examination of the multifaceted moral and ethical considerations that cannot be fully accounted for with the use of quantitative approaches.

Using a comparative framework, methodology consists of analyzing traditional human driven social engineering with evolving AI inspired methods. The two approaches are assessed in the light of [4]Joseph M. Hatfield's virtue ethics framework, which centres on moral qualities of responsibility, moderation, and the weight put on communal well being, to ascertain the ethical permissibility of each. This comparative analysis is conducted in two main phases:

Literature-Based Analysis: In this phase, we review existing ethical frameworks, such as utilitarianism and deontology, so as to provide context, as well as a contrast with the application of virtue ethics in cybersecurity.

Framework Application: In this, virtue ethics in specifically applied to case studies of traditional and AI based social engineering practice, towards understanding the ethical dimensions such as moderation, accountability, and wider organizational impact upon the community.

3.2 Data Collection Process

3.2.1 Literature Review and Theoretical Foundations

The literature review included extensive collection of data from a variety of sources focused on ethical frameworks relevant to cybersecurity and social engineering in order to get started. In this phase, the key philosophical perspectives of virtue ethics, utilitarianism, and deontology, were reviewed from the perspectives of how each articulated their view of ethical boundaries of social engineering practices. Along with the recent studies that tackle the unique difficulties in using AI and ASE for social engineering (e.g. scalability, impersonality, lack of accountability if AI acts independently, etc.) are considered by the literature review.

This theoretical foundation places social engineering practices within these general ethical discussions and brings to light how virtue ethics, with its full weight on the states of mind and its connection to community, is different from outcome focused or rule based ethics. Analyzing both traditional and AI driven social engineering practice in a penetration test setting is only possible with this understanding.

3.2.2 Case Study Selection

The study developed two hypothetical case studies to explore and contextualize ethical issues in social engineering:

Traditional Social Engineering Case: In this case, a phishing email aimed at the employees within an organization is used to leverage a human discretion and adaptability. Through this case we are able to examine ethical boundaries within traditional penetration tests where a human is involved in each interaction.

AI-Driven Social Engineering Case: In this scenario, an AI powered phishing campaign at scale, hitting several people at once. It is a fully autonomous AI running with minimum to no real time human intervention, and therefore raises ethical question related to accountability, possible harm, and also a balance between protecting an organization and personal autonomy.

The application of the virtue ethics framework through these case studies makes these case studies be controlled scenarios of applying virtue ethics structured principles to assessing the appropriateness of using social engineering methods in different contexts. In seeking to demonstrate the practical ethical challenges in actual world conditions especially where human judgment and AI automation meet in the world of penetration testing, the selected case studies aim to reveal.

3.3 Tools and Techniques for Analysis

3.3.1 Framework Application and Ethical Analysis

To better understand ethical decision making criteria in social engineering practices, the above case studies are approached within the virtue ethics framework. This framework assesses three core dimensions:

Moderation: It evaluates the degree to which social engineering practices find an ethical middle ground between the requirement of organizational security and individual rights. This dimension looks at the implications of balance in favor of traditional versus AI driven methods with the question of whether some tactics may overstep ethical boundaries through over manipulation.

Responsibility: Fully examines social engineering accountability, examining the distinction between classical human powered and artificial intelligence driven consequences, on their ability to minimize harm and maintain ethical obligation. As opposed to AI driven practices that might not have real time realibility of ethical change, traditional methods with human oversight inherently extends the possibility of real time ethical change.

Community Impact: It studies the broader implication of social engineering to organizational as well as communal welfare. This dimension reflects the trade off between the security organization gains and the losses of employee trust and autonomy, in particular in as far as work environments involving AI tend to involve the impersonal enterprise of automation that generates dissonance within the organizational community.

3.3.2 Comparative Ethical Analysis

A comparative ethical analysis is also employed to evaluate traditional and AI-driven social engineering practices across three ethical frameworks: However, it looks at virtue ethics, utilitarianism, and deontology. With each separate framework you have different starting point to immerse yourself in the view, as you have the holistic picture from a social engineering standpoint of what is moral and what is morally not permissible.

Virtue Ethics: In a way it is tempered in the moral character of actions, so to speak, it offers a latitude to adjust the ethical part according to particular circumstances. This approach dictates that for an organization and individuals' fortune and the well being of some common moral values must be considered when making ethical decisions.

Utilitarianism: The social engineering tactics is examined in terms of their applicability to the practical results, valuing the actions which deliver the most good and least damage. However, an output quantity driven approach to such an outcome can ignore rights, which is important when large scale ASE creates unintended consequences.

Deontological Ethics: There are rules therefore that emphasize how important it is to follow them and how important the prohibition in a categorical way appeared to deceive and socially manipulate. For example, this framework claims that such social engineering isn't permitted, requiring transparency and consent, but some types of penetration testing may or may not apply.

3.4 Ethical Considerations

The study was conducted in strict adherence to ethical guidelines, ensuring no real data or identifiable personal information was collected or analyzed. Ethical considerations in this research

prioritize the minimization of harm and the protection of individual rights while supporting organizational security needs. By utilizing hypothetical case studies, the research maintains ethical integrity and academic rigor, facilitating a balanced examination of social engineering tactics without infringing on individual privacy or autonomy. These safeguards reflect a commitment to both ethical standards and the academic objectives of understanding ethical issues in cybersecurity.

4 Analysis and Discussion

4.1 Analysis of Research Findings

4.1.1 Ethical Considerations in Traditional Human-Driven Social Engineering

Based on this virtue ethics framework presented in the methodology, the analysis of existing traditional human driven social engineering shows how human oversight provides for the common flexibility of real time ethical adjustments[4](Hatfield, 2020). However, humans play a crucial role in application of ethical principles and they present a practical and balanced model also in traditional forms of social engineering practices where testers can use their discretion to see how people react and adjust yours approach in a way to avoid harm. This corresponds well with the virtue ethical literature review – the ethical concepts of virtue ethics of moderation and responsibility.

Virtue ethics offers a way, however, for human-driven social engineering to navigate such complex situations in an ethically responsible way — at just the level of manipulation necessary to respect individual autonomy while securing organizational goals. According to Hatfield's framework, the social engineering should have moderated use of manipulation in their conduct to achieve communal good rather than extremes. The outcome of this analysis validates that social engineering practices can be ethically sound when driven by human discretion: that it is possible to tame social engineering practices in ways that respect individual rights and promote community values, which is exactly what virtue ethics also suggests —with a strong wager on moral responsibility.

4.1.2 Ethical Challenges in AI-Driven Automated Social Engineering (ASE)

This is in contrast to AI driven ASE, where there are unique ethical challenges arising from a lack of real time human oversight, and the ability for ethical discretion. Unlike its conventional counterparts, AI based systems do big scale automated social engineering without self-adjusting ethics on an individual level. The depersonalization of crime causes additional ethical worry since a lack of situational awareness increases the risk for both the individual and larger scale harm. According to the literature review, virtue ethics relies on moderation and accountability, two principles that become difficult to maintain when operating with AI driven processes.

What they find is that automation minus human judgment is a poor fit for the ethical principles of moderation and community good. However, AI driven tactics are based on the pre defined algorithms and are lacking in flexibility to run ethically on unforeseen cases. However, this disconnect underscores a fundamental failing of virtue ethics when applied to AI systems, since virtue ethics presumptions on a human agent to act and morally. As a result, AI driven ASE results in efficiency and scalability but at the expense of ethical considerations, requiring adapted guidelines for automated contexts.[1](Floridi & Cowls, 2012)

4.2 Interpretation of Results

Ethical adaptability and real-time accountability are the aspects of stark contrast between human driven and AI driven social engineering. First and foremost, in traditional penetration testing, each interaction is led by human discretion, allowing for dynamic responses as well as ethical principles to guide that interaction. While in that form, virtue ethics can be applied in situations where rigidity of automation restricts the ability for moral discretion, this kind of rigidity presents challenges for moral discretion in AI driven scenarios.

The results demonstrate that although virtue ethics has a strong ethical foundation for social engineering seeking human participation, further thought is necessary for application to ASE. Given the lack of human oversight in real time in AI driven process, there is need for evolving ethical framework so as to address operational constraints and the scalability to automated system. In other words, virtues ethics to make sense of AI contexts must be combined with AI ethics principles based on transparency, harm minimization and responsibility in autonomous actions.

4.3 Discussion of Findings in Relation to Research Objectives

The findings meet directly the objectives of the study on the applicability of virtue ethics principles to attend the social engineering from traditional and automated point of view. This was one of the main objectives, in order to evaluate how to guide penetration testing with virtue ethics when AI driven ASE introduces new ethical complexities. The analysis in this chapter shows that although social engineering provides valuable ethical wisdom for traditional social engineering, virtue ethics is not effective in ASE contexts where human adaptability and judgment are lacking.

The scope of this limitation suggests that AI driven systems call for a reformulation of existing ethical frameworks to address the challenges at hand. The research objectives have identified these gaps as essential in filling the gap to develop a framework to support the integrity of ethics in cybersecurity which is at its infancy. The results highlight the importance of ethical responsibility and moderation in organizational cohesion and maintaining trust for individual autonomy in security testing practice.

4.4 Implications for Future Research and Practical Application

Future research is suggested in which virtue ethics principles are applied to ASE but with AI specific ethical principles integrated in traditional frameworks. In future studies it could be investigated how AI ethics can be integrated with guidelines, whichpertain to transparency, responsibility and harm reduction in the context of autonomous systems. Additionally, cybersecurity practitioners might incorporate in AI driven penetration testing oversight mechanisms so that ethical standards are adhered to even where human judgement does not directly engage. This approach could bridge the gap between virtue ethics and the practical demands of AI driven cybersecurity practices and therefore support ethical conduct.

5 Case Studies

5.1 Traditional Human-Driven Social Engineering Case Study

Scenario: A large financial institution runs a penetration test using traditional social engineering techniques to test their employee's vulnerability to phishing attacks; a cybersecurity team at this same institution used the same docs to run social engineering attacks as part of an assessment of their user base. For example, in this case penetration tester sends a phishing email in which the sender portrays himself to be a member of the IT department who needs employees to reset their passwords because of a 'security update.' Responses are watched in real-time by the tester, adjusting their tactics according to how an individual reacts. As an example, if the employee is confused in their request, the tester follows up to understand if an internal communication request might be made and to assess the level of trust employees have in those communications.

Ethical Analysis: In this case study, we make an observation about the ethical flexibility of human-driven social engineering from the perspective of virtue ethics. The moderation, a central principle of virtue ethics, comes from the tester, who can react differently to each of those reactions. The dynamic nature of the responses means that while the tester can manipulate and which way, the tester can also safely caliper the level of manipulation to minimize potential psychological damage by tracking their interaction, respecting need for autonomy while simultaneously working towards the needs of the organization.

Closely reflected in virtue ethics framework provided by Joseph M. Hatfield, this stand out as a responsibility and a common good case. The tester's decisions are also for the benefit of everyone because they locate security vulnerabilities that threaten the institution but otherwise do not infringe on the rights of employees by using overly provocative methods. This stands well with the research objective of exploring how virtue ethics can help human centred social engineering for security and rights of the organization and the person.

This case shows that employing virtue ethics in practice makes the classic approach to social engineering ethical as well as secure and, at the same time, lets people choose for themselves. This shows that human interaction is required especially because decision making is situational when it comes to ethical decision making during penetration test.

5.2 AI-Driven Automated Social Engineering (ASE) Case Study

Scenario: A social engineering system incorporated with Artificial Intelligence is implemented to launch a comprehensive phishing attack in the different departments of a large-scale, multinational company. The AI system, which mimics the official correspondence of a corporation, at once sends thousands of phishing messages saying there is a 'mandatory cybersecurity training update' where the employees should click the link. It is an automated system and lacks the ability to realize the actual reactions and affect the given campaign if negative impact occurs.

Ethical Analysis: The following is a real life example of how ethicality is restricted concerning the use of AI-ASE in virtue ethics such as moderation and responsibility. However, to be fair, the AI system under discussion cannot be as flexible and discriminate in real-time operations as a human counterpart is within the framework of virtue ethics in social engineering. In addition, the AI is not able to assess the distinct responses of the audiences or pause the campaign, and as a result, the decisions it makes may cause perplexity and doubt, as well as even bring about

psychological harm, which violates the ethical principles, and the purpose of virtue ethics.

In this case such an arrangement provides the AI complete discretion in this specific instance, and of course such a decision raises a lot of ethical concerns given that virtue ethics draws its roots from the idea of a community. For example, it lacks individual obligation different to human-reliant social engineering subclass which shows that the subclass can warp ethical too. The experience in agency captured by the case of the flawed AI system is that such a system cannot be selective and safeguard communal interest while altering its processes to evade risk. This is in harmony with under Analysis and Discussion session, for instance, the failure of utilizing virtue ethic in AS, it cannot make moral decisions nor transform positively for the benefit of the society.

From this case study, it is clear that there is a requirement for proper ethical standards for Artificial Intelligence operating networking systems in social engineering. It supports the research's conclusion that virtue ethics, despite being efficient in the conventional work settings, needs enhancement in concerns to the execution limitations of AI-integrated ASE. Such a scenario is also relevant to the research aim of mapping blind spots in currently available ethical paradigms, as the AI-driven social engineering systems suggest the necessity of the use of oversee mechanisms to ensure the ethical approach to cybersecurity.

5.3 Comparative Analysis of Case Studies

Altogether, all these case studies point towards how valuable and indispensable human judgement and decision-making are in the practical application of virtue ethics to social engineering. The free response testing and judgment capacity for the tester help this case to come closer to a blend the virtue ethics that balance security for the organization and that of the user. However, AI-dominated situation shows that ethically AI cannot manage actions in real time and cannot take moral responsibility.

In sum, all these examples demonstrate the general picture of ethical-pluralist approach in cybersecurity. They emphasize the suggestion that although the virtue ethics may offer a sound ethical undercurrent for the conventional social engineering, it has its drawbacks in the ASE context, which is devoid of human supervising. These results support the study recommendation of understanding the proliferation of ethical frameworks that include the use of AI ethics principles in ensuring the social engineering process remains ethical by communicating accountability and reducing harm in autonomous systems for both the human led and AI led sub-processes.

6 Recommendations

6.1 Practical Recommendations for Cybersecurity Practice

6.1.1 Implement Oversight Mechanisms for AI-Driven Social Engineering

Due to the inherent ethicality of ethical discretion in AI-driven automated social engineering (ASE), it is recommended that organizations establish monitors to check the ethicality of decisions made where a human can make an ethical decision at the time. These mechanisms could include:

Human-in-the-Loop Monitoring: Introducing periodic human oversight into ASE campaigns to allow the campaigner to brief that is, the researcher to intervene if unethical effects are apparent

in real-time. This is in harmony with the virtue ethics of moderation as well as the accountability which permits alterations that will decrease the effect of injury.

Ethical Review Panels: Formulate internal committees that evaluate the characteristics and possible effects of the AI driven campaigns before they go live. This means strategies and messages stay on the right side of the law and organization code of conduct which provides with Hatfield's guideline of safeguarding community interest.

These recommendations restore certain amount of human responsibility to make ethical decisions, mentioning the lack of it in AI-based social manipulation. They address directly, concepts in the Analysis and Discussion and Case Studies whereby the need for flexible supervision in automated environments was deemed necessary.

6.1.2 Develop AI-Specific Ethical Guidelines

As was described, traditional or classical approaches for ethical guidelines proven sufficient for human orchestration of social engineering, do not prove enough for AI-led social engineering. Thus, to respond to the ethical requirements inherent in ASE, some guidelines concerning virtue and AI ethics should be adopted. Core elements of these guidelines should include:

Transparency Requirements: Set, specify and standardize available communication structures for ASE campaigns that shall be embraced within the organization, to avoid doubts that may hinder ethical practices in the organization.

Limitations on Scope and Scale: Likewise, set limits regarding the size and type of targeting to be used in AI-driven campaigns so that AI cannot over-push and control people adversely.

Assign Clear Accountability: Assign particular changes to particular individuals or departments to assume responsibility to guarantee that someone is overservice even when the transmission is computerized.

These AI-specific guidelines present an ethical scaffold that lies within the tradition of virtue ethics but answers to the matter of ASE and its problematic issues associated with large scale subjugation and deindividuation. This presume from the Methodology and Case Studies, that there is a demand for an ethical prudery which common virtue ethic cannot deliver in automated environments.

6.1.3 Integrate Ethical Training Programs for Cybersecurity Professionals

This paper argues that ethical training of penetration testers and cybersecurity professionals is crucial to maintaining order in the application of virtue ethics in social engineering. Training should include:

Ethical Frameworks Overview: Covers all aspects of virtue ethics, utilitarianism, and deontology, to help the Practitioner understand how ethical concepts are applied in various fields of social engineering.

AI Ethics Modules: They should comprise specific training on the ASE specific ethic concerns like harm minimization and operational accountability.

Scenario-Based Learning: Discussed here hypothetical examples similar to those of the Case Studies that would help practitioners get real-world experience about addressing ethical issues in both possible human and AI facilitated social engineering.

This recommendation corresponds to the Literature Review and Analysis and Discussion where flexibility and decision-making components known as the core of the generally accepted social engineering strategies were identified as critical for the ethical standards. This way, through development of knowledge to handle ethical dilemmas in the professional setting, these organizations can enhance ethical standards both in manual and A.I. models utilized in presenting them.

6.2 Suggestions for Future Research

6.2.1 Development of Hybrid Ethical Frameworks for AI-Driven Social Engineering

The analysis showed that, besides, there are issues with the traditional ethical theories, especially, with virtue ethics theory when it would be applied to the AI-driven social engineering. Similarly, more studies should be carried out with the aim of findings more versatile models founded on virtue ethics as well as AI ethic principles such as disclosure, non maleficence and accountability. I would propose such framework to offer the ethical ambiguity and definite character for cybersecurity specialists to meet the moral challenges of ASE.

This suggestion corresponds with the points made in Analysis and Discussion parts of the paper, where lack of applicability of virtue ethics in automated situations was discussed. Such gaps could be bridged by hybrid frameworks in order to enhance the general use of ethical AI in cybersecurity.

6.2.2 Investigate Long-Term Psychological Effects of ASE on Employees

There is little known about the extent of employees exposed to AI driven socially engineered attacks, especially in large scale scams. Possible research questions for future research could include examining the long term effect that ASE has had on employees' level of trust, morale and the organisational culture. The data offered in such studies would contribute towards defining norms where the people's effectivity comes first while not underestimating the work security processes.

This is based on the outcomes in Case Studies which have reported that people felt depersonalized and possibly other psychological consequences in event of the operation of artificial intelligence. The idea related to the long-term consequences of ASE, will also foster good practice for security without negatively impacting staff.

6.2.3 Examine Legal and Regulatory Implications of AI in Social Engineering

The legal and regulatory aspects of the AI application in cybersecurity remain undefined and require further study of responsibility in automated social engineering. More precisely, this research should find out how present legal instruments regulate practices involving AI and put forward recommendations for ethical ASE adoption. The results might help in the formulation of measures that promote ethical behaviour while staying legal.

This recommendation is thereby based on the Analysis and Discussion which had identified accountability problematics in AI based systems. Legal research would offer organizations how to unravel these issues with an ethical and legal approach.

7 Conclusion

This research aimed at enhancing understanding of ethical issues within social engineering in cybersecurity field, particularly comparing the use of conventional integrative manual social engineering to automated social engineering (ASE) within penetration testing environment. Using virtue ethics as a primary lens, in conjunction with utilitarian and deontological perspective, the study assessed how the approaches meet organizational security requirements while respecting the autonomy of individuals. The literature review has highlighted the role of ethical frameworks in cybersecurity more specifically where the existing ethical frames are seen not to fully apply when implemented to automated systems. The qualitative approach using hypothetical case studies shown in this research and this research showed that human-driven SE has advantages in ethical flexibility on human control and monitor due to the human in-time changing of strategies according to people's reaction . This flexibility complies with the virtue ethic principles for moderation and responsibility that promote proportionality between organizational security of data and users' rights. However, the assessment revealed that although adopting AI for ASE is significant in scale, it is fundamentally devoid of the ethical decision-making variability and responsiveness needed by virtue ethics. Automated systems lack the discretion not to harm the individual and the decision making leading to such ethical issues as depersonalizing, lack of individual accountability, potential harm to numerous people. Hence there is a demand of a dual ethical approach where one is adapted to tackle the ordinary ethical concerns while the other is unique to the world of AI and dealing with ASE. Recommendations for applying cybersecurity practice involve Such recommendations include: including measures of oversight; establishing best ethical standard for specific practice; and incorporating ethical training to support the practitioner to gain knowledge on how tackle ethical dilemmas in both human manipulated and artificial intelligence manipulated social engineering[2](Kshetri, 2019). This study also exposes the drawback of the current prevailing ethical paradigms and gives future research a point of departure for continuing to progress the field of AI ethics in cybersecurity and guarantee that advanced technologies in this area are secure while also being developed under the principles of ethical status. Looking back at the research approach, the study highlighted the need for ethical frameworks that are equally dynamic to meet existing and emerging technological development in a way that protects organizational goals and individual subjectivity in a future defined by AI.

8 Refrence

References

[1] Luciano Floridi and Josh Cowls. "A unified framework of five principles for AI in society". In: *Machine learning and the city: Applications in architecture and urban design* (2022), pp. 535–545.

[2] Naresh Kshetri. *The global rise of online devices, cyber crime and cyber defense: Enhancing ethical actions, counter measures, cyber strategy, and approaches*. University of Missouri-Saint Louis, 2022.

[3] Mark Coeckelbergh. "Mark Coeckelbergh: AI Ethics". In: (2021).

[4] Joseph M Hatfield. "Virtuous human hacking: The ethics of social engineering in penetration-testing". In: *Computers & Security* 83 (2019), pp. 354–366.

[5] Luciano Floridi and Mariarosaria Taddeo. *What is data ethics?* 2016.

[6] Francois Mouton et al. "Social engineering attack framework". In: *2014 Information Security for South Africa*. IEEE. 2014, pp. 1–9.

[7] Peter Finn and Markus Jakobsson. "Designing ethical phishing experiments". In: *IEEE Technology and Society Magazine* 26.1 (2007), pp. 46–58.

[8] Mary Gregor et al. *Groundwork of the Metaphysics of Morals*. Cambridge University Press Cambridge, 1997.

[9] T Aristotle and M Ostwald. "Nicomachean ethics, book VIII". In: *Bobbs-Merrill Co Indianapolis, IN* (1962).

[10] Jeremy Bentham. "An introduction to the principles of morals and legislation". In: *History of Economic Thought Books* (1781).