Issues in Information Systems
Volume 26, Issue 2, pp. 452-465, 2025

DOI: https://doi.org/10.48009/2_iis_134

Artificial intelligence in social engineering: a literature review through the lens of routine activity theory

Chloe Dzuba, *Robert Morris University, ccdst263@mail.rmu.edu*

Abstract

Social engineering attacks have become more sophisticated with the emergence of artificial intelligence (AI), enabling cybercriminals to automate and scale their attacks. This paper examines the intersection of AI and Routine Activity Theory (RAT) to understand how AI-driven social engineering challenges traditional criminological frameworks. By analyzing AI-enhanced phishing, deepfakes, and automated interactions, the study explores how RAT's core elements—motivated offenders, suitable targets, and the absence of capable guardians—apply to modern cyber threats. A comparative analysis of traditional and AI-enhanced attacks is conducted to evaluate the evolving dynamics of social engineering. The paper argues that RAT, while foundational, must evolve to address the complexities introduced by AI, and proposes adaptations to improve its applicability in the context of emerging cybercrimes. The findings suggest that interdisciplinary approaches combining AI, criminology, and cybersecurity are essential to developing effective prevention and mitigation strategies in the digital age.

Keywords: artificial intelligence (AI), criminology, social engineering, routine activity theory (RAT), AI-enhanced social engineering, cybersecurity

Introduction

Social engineering attacks have become increasingly sophisticated in the ever-evolving cybersecurity landscape, leveraging new technologies to deceive and exploit individuals. Traditionally, social engineering attacks relied on human interaction and psychological manipulation to gain access to sensitive information or systems (Hatfield, 2018). However, with the advent of artificial intelligence (AI), these attacks are dramatically transforming, becoming more automated, scalable, and complex (Blessing et al., 2022). AI technologies such as machine learning, deep learning, natural language processing, and automation are revolutionizing social engineering techniques, enabling attackers to craft highly personalized and convincing phishing schemes, deepfake content, and even automated voice and chat interactions. This paper seeks to explore the role of Routine Activity Theory (RAT) in understanding the dynamics of AI-powered social engineering attacks, evaluating whether this criminological framework can adapt to the changes in the methods and tools employed by cybercriminals.

Routine Activity Theory, first proposed by Lawrence Cohen and Marcus Felson in 1979, has served as a foundational framework in criminology, providing insight into the conditions that allow crime to occur. The theory hinges on the interplay of three critical elements: a motivated offender, a suitable target, and the absence of a capable guardian. According to RAT, criminal events are not solely the result of individual motivation but are significantly influenced by the routine activity of individuals and the context in which they operate (Cohen & Felson, 1979). Traditionally applied to physical crimes, RAT has also been adapted

to understand the dynamics of cybercrime, as the internet and digital technologies have created new avenues for criminal activity (Ahmand & Thurasamy, 2022). In the context of cybercrime, RAT argues that online offenders—often anonymous and globally distributed—can easily identify vulnerable targets in digital spaces where guardianship is typically minimal or non-existent (Holt & Bossler, 2008). However, the rapid advancement of AI in recent years has presented significant challenges to this framework, particularly in the context of social engineering attacks. As AI becomes increasingly integrated into cybercriminal activity, its capabilities introduce significant changes in how crimes are committed, thus raising critical questions about whether RAT can still adequately explain these modern threats (Ahmand & Thurasamy, 2022).

The rise of AI in social engineering presents new challenges and opportunities for both attackers and defenders. On the one hand, AI can enhance the capabilities of cybercriminals, allowing them to automate and optimize attacks, reach a larger number of potential victims, and bypass traditional security measures (Alahmed et al., 2024). On the other hand, the widespread adoption of AI also means that defenders must adapt their strategies to keep pace with these emerging threats (Blauth et al., 2022). The convergence of AI and RAT raises several pertinent questions: Does RAT still apply to AI-driven social engineering attacks? How does AI impact the routine activities of potential victims, and what role do AI-driven technologies play in the absence of capable guardianship? How might criminological theories evolve to better explain and predict AI-enhanced social engineering tactics?

This paper will explore these questions through a content analysis of current social engineering attacks, comparing AI-driven incidents with traditional social engineering cases. By analyzing real-world examples, reports, and case studies, this paper will categorize attacks based on the elements of Routine Activity Theory and assess how the framework can be applied to both traditional and AI-driven social engineering crimes. Through this analysis, the paper will also examine whether current criminological theories, including RAT, can evolve to address the changing dynamics of cybercrime in the age of AI.

This paper will also provide an overview of Routine Activity Theory and its application to cybercrime. It will then explore the role of AI in transforming social engineering techniques, focusing on AI-enhanced phishing, deepfakes, and automation. A comparative analysis of traditional and AI-driven attacks will follow, highlighting the similarities and differences in how these crimes are carried out and the implications for RAT. Finally, the paper will propose potential adaptations or modifications to RAT that would allow the theory to better capture the complexities of AI-powered social engineering attacks and provide insights into how law enforcement, security professionals, and policymakers can respond to this emerging threat.

As cybercrime continues to evolve with the rise of AI, understanding the relationship between criminological theories like RAT and new forms of social engineering is crucial for developing effective prevention and mitigation strategies. By reexamining RAT in the context of AI, this paper aims to contribute to the ongoing discourse on how criminological frameworks can adapt to the changing nature of cybercrime and provide valuable insights into the future of cybersecurity.

Literature Review

Routine Activity Theory (RAT) was originally developed by Cohen and Felson (1979) and explains crime occurrence through the interaction of three key elements: a motivated offender, a suitable target, and the absence of a capable guardian. Traditionally applied to physical crimes, RAT has been adapted to cybercrime (Cyber-RAT), where digital environments create new opportunities for victimization. Leukfeldt and Yar (2016) argue that RAT can't definitively be found to be suitable or unsuitable in applicability to cybercrime.

Holt and Bossler (2008) examined RAT's applicability to cybercrime victimization, emphasizing that users' online routine activities, such as a lack of cybersecurity awareness or frequent social media engagement, influence their likelihood of being targeted. Ahmad and Thurasamy (2022) conducted a systematic literature review on RAT in cybercrime, highlighting gaps in the theory when applied to advanced cyber threats. They noted that while RAT is commonly used to understand factors that influence cybercrime victimization, there are inconsistent findings on its applicability.

While RAT provides a solid foundation for analyzing cybercrimes, the emergence of artificial intelligence (AI) in social engineering attacks challenges its core assumptions. AI tools enable attackers to operate with unprecedented efficiency, automating deception tactics and reconnaissance (Schmitt & Flechais, 2023). This raises questions about whether "motivated offenders" still fit the traditional human-centric model or if AI itself functions as an autonomous offender (Blauth et al., 2022). Some researchers argue that AI's role in cybercrime blurs the line between offender and tool, necessitating a deeper examination of RAT's applicability in AI-driven cybercrime (Leukfeldt & Yar, 2016).

Social engineering attacks, such as phishing and impersonation scams, rely on psychological manipulation to deceive victims into divulging sensitive information or performing harmful actions (Alahmed et al., 2024). Traditionally, these attacks were carried out manually, requiring human effort to conduct reconnaissance, craft persuasive messages, and engage with their targets (Hatfield, 2018). However, advancements in AI have drastically transformed the landscape of social engineering, enabling cybercriminals to automate and scale their attacks with unprecedented precision (Blauth et al., 2022). AI-driven tools allow attackers to tailor messages to specific individuals or organizations, increasing their effectiveness by exploiting personal details, behavioral patterns, and communication styles.

Schmitt and Flechais (2023) highlight how generative AI tools, such as AI-powered chatbots and deepfake technology, have introduced sophisticated attack vectors that blur the line between legitimate and fraudulent communications. AI-generated phishing emails now exhibit flawless grammar, context-aware phrasing, and persuasive language, making them far more challenging to detect than traditional scams. Additionally, deepfake technology allows attackers to create hyper-realistic voice messages and video impersonations, significantly increasing the success rate of business email compromise (BEC) scams and executive fraud. As organizations implement security training programs emphasizing traditional red flags, AI-enhanced deception techniques continue to outpace conventional detection methods.

Similarly, Alahmed et al. (2024) examine the role of AI in misinformation campaigns and phishing, emphasizing how generative AI eliminates common indicators of fraud. Previously, users could often identify phishing attempts based on poor sentence structure, grammatical inconsistencies, or generic messaging. However, AI-generated content can mimic the writing style of specific individuals, predict user responses, and personalize interactions based on historical data, making it significantly more challenging for even security-conscious individuals to recognize threats. Furthermore, AI-driven social engineering is not limited to emails— attackers now use AI-generated phone calls, chatbots, and social media interactions to manipulate victims into disclosing credentials or approving fraudulent transactions.

These developments illustrate a growing challenge in cybersecurity, as AI-powered attacks continue to evolve faster than defensive measures. The ability of AI to analyze vast amounts of personal data, generate highly targeted messages, and simulate realistic human interactions presents a significant risk to both individuals and organizations. As a result, traditional social engineering countermeasures, such as employee training and email filtering systems, must be augmented with AI-driven defense mechanisms capable of detecting subtle anomalies in communication patterns. Addressing these issues will require a combination

of advanced AI-based threat detection, continuous security awareness training, and proactive policy enforcement to mitigate the risks posed by increasingly sophisticated AI-generated social engineering attacks.

Blauth, Gstrein, and Zwitter (2022) further explore the malicious use of AI in cybercrime, noting that AI can conduct reconnaissance on potential targets by analyzing leaked datasets, social media activity, and publicly available information. This automated target selection process reduces the time and effort required for attackers to find vulnerable individuals, making AI-driven social engineering a scalable and low-cost threat. With AI performing much of the attacker's workload, traditional assumptions about human effort in cybercrime need to be reassessed.

The integration of AI into social engineering raises several challenges for Routine Activity Theory. It modifies key components of the theory in ways that have not been fully addressed in criminological research. Historically, RAT assumes that a human offender is actively engaging in criminal behavior. However, the rise of AI-driven social engineering challenges this assumption, as machine learning and natural language processing (NLP) now enable the automation of highly convincing phishing techniques (Blauth et al., 2022). Generative AI models, such as large language models (LLMs), further blur the line between human and machine deception by crafting realistic dialogues. This has led to the emergence of using tools like ChatGPT to create targeted, persuasive messages with minimal effort (Gupta et al., 2023).

Broadhurst et al. (2019) and Hayward and Mass (2020) argue that AI changes this paradigm by enabling "offense as a service," where cybercriminals leverage AI tools to commit cybercrimes more efficiently. In some cases, AI may function as an autonomous offender, generating and deploying social engineering attacks with minimal human input. This raises a critical question: Can AI be classified as a "motivated offender" under RAT, or is it merely an extension of human criminal intent?

Furthermore, AI-driven attacks complicate traditional concepts of capable guardianship. Conventional cybersecurity measures, including antivirus software, firewalls, and awareness training, are increasingly ineffective against AI-enhanced threats that dynamically adapt to evade detection. Jeong (2020) highlights that these attacks mimic human behavior, making them difficult to distinguish from legitimate interactions. This challenges the ability of law enforcement and security professionals to respond effectively. As AI continues to evolve, its role in cybercrime necessitates a reevaluation of RAT's core assumptions to address emerging threats in the digital landscape. Additionally, AI-generated synthetic identities pose a challenge to existing authentication and verification processes. Unlike human-generated phishing emails that existing filters can often flag, AI-crafted messages dynamically evolve to bypass security measures (Schmitt & Flechais, 2023). This dynamic adaptive nature of AI-driven crime necessitates a reevaluation of what constitutes capable guardianship of digital spaces.

With AI automating reconnaissance and target selection, the range of potential victims has significantly expanded. Unlike traditional cybercriminals who may manually select high-value targets, AI can identify and exploit vulnerable individuals or organizations at scale. This reduces the cost and effort required for attacks, making even small-scale fraud operations more profitable. As a result, RAT's traditional notion of a "suitable target" must account for the mass scalability of AI-driven attacks, which challenge the theory's assumptions about offender effort and target selection.

Despite the threats posed by AI-driven social engineering, AI can also be leveraged as a tool for mitigation and crime prevention. Several researchers propose AI-driven solutions to counteract emerging threats. AI-Powered Threat Detection: Broadhurst et al. (2019) and Blessing et al. (2022) discuss how AI-driven security tools can analyze communication patterns, identify fraudulent attempts, and detect anomalies in

real time. These systems use machine learning algorithms to continuously improve their accuracy in detecting AI-generated attacks and threats. Regulatory and Legal Responses: Mohsin (2020) highlights the importance of AI governance and legal frameworks to criminalize AI-generated deception and regulate its use. Without proper oversight, criminals will continue to exploit gaps in AI regulation to advance their methods. Human-Centered Cybersecurity Awareness: Kayser, Mastrorilli, and Cadigan (2019) emphasize the role of human vulnerabilities in cybercrime victimization and advocate for comprehensive cybersecurity training that includes awareness of AI-generated threats. Educating users on recognizing AI-enhanced social engineering tactics is critical in strengthening human guardianship.

While Routine Activity Theory has been influential in explaining cybercrime, AI-driven social engineering challenges its traditional framework. The evolving nature of AI-driven cybercrime raises several key questions: Should AI itself be considered an offender, or is it merely a tool in the hands of human criminals? How can the concept of capable guardianship be expanded to account for AI-driven threats? Does the suitability of targets change when AI automates the selection and exploitation of victims?

To analyze how AI enhances social engineering attacks applied to Routine Activity Theory, I will be examining how AI influences each stage of the social engineering process, with new complexities introduced that require further theoretical adaptation.

AI-Assisted Attacker: AI lowers the barrier to entry for cybercriminals, automating tasks such as reconnaissance, deception, and exploitation. Tools like deep learning algorithms and natural language processing (NLP) enhance impersonation, phishing, and social manipulation techniques (Blessing et al., 2022). Attackers can craft compelling messages, create synthetic identities, and adapt attacks in real time based on victim responses (Gupta et al., 2023).

Lack of Capable Guardianship: Traditional defenses struggle against AI-powered attacks. Machine learning models can evade detection by modifying malicious content to bypass spam filters and/or anomaly-based detection systems (Schmitt & Flechais, 2023). Deepfake technology further weakens digital authentication, hindering conventional security measures. This lack of effective countermeasures creates new vulnerabilities in online spaces.

Suitable Targets: AI enhances target selection by analyzing large datasets to identify individuals who are susceptible to manipulation tactics. Behavioral profiling and social media scraping allow attackers to personalize scams, increasing their success rate (Kayser et al., 2019). High-profile individuals and organizations are particularly vulnerable due to their digital exposure.

Reconnaissance: AI-driven Open Source Intelligence (OSINT) tools automate the information-gathering process, identifying vulnerabilities more efficiently than human attackers. Machine learning models analyze leaked credentials, communication patterns, and job postings to map potential access points (Alahmed et al., 2024). This enables AI-powered attackers to craft hyper-personalized attacks with minimal effort.

Deception: The sophistication of AI-generated deception surpasses traditional methods. AI-generated text, deepfake audio and video, and chatbot-driven scams make distinguishing real from fake increasingly difficult (Caldwell et al., 2020). Attackers exploit this technology to impersonate executives, manipulate victims into revealing sensitive data, and conduct business email compromise (BEC) scams.

Attack Vectors: AI refines existing attack vectors, making traditional phishing, vishing (voice phishing), and smishing (SMS phishing) more dangerous. Chatbots and AI-driven social media accounts can automate and scale social engineering attacks, making them more effective and persistent. (Blauth et al., 2022).

Exploitation: AI tools dynamically adjust manipulation tactics based on victim behavior. Adaptive persuasion techniques allow attackers to increase the likelihood of success by analyzing emotional cues and linguistic patterns in real time (Caldwell et al., 2020).

Crime: AI-driven social engineering challenges criminological and legal frameworks. The scale, speed, and sophistication of these attacks blur traditional distinctions between cybercrime and automated fraud. This study explores whether RAT requires theoretical modifications to account for AI's role in crime commission and how security professionals, policymakers, and law enforcement can respond.

Future research should explore whether Routine Activity Theory needs to be expanded to incorporate AI as an independent factor influencing cybercrime dynamics. Additionally, interdisciplinary studies combining AI ethics, criminology, and cybersecurity will be crucial in developing comprehensive strategies for mitigating AI-driven social engineering threats.

Methodology

In this research, a content analysis methodology was used to analyze existing reports systematically, news articles, case studies, and cybersecurity incident databases. The primary goal was to categorize instances of social engineering attacks, both traditional and AI-enhanced, based on the elements of Routine Activity Theory (RAT) while applying five derived parameters that map the progression of AI-driven social engineering: reconnaissance, deception, attack vectors, exploitation, and crime. These parameters were synthesized from recurring themes both in conventional attacks and emerging AI-driven threats.

Data was collected from a wide range of reliable sources, including institutions like Proofpoint, Microsoft Digital Defense Report, Onfido's Identity Fraud Report, Mimecast Global Threat Intelligence Report, CrowdStrike, and The Hacker News. This selection focused on social engineering, AI-powered attacks, and crime, emphasizing incidents from the last five years to reflect currecncy. Sources included both quantitative and qualitative data, such as attack types, attack success rates, or victim demographics.

Each case study was analyzed using a structured thematic approach, with a two-tiered coding system. The first tier of Routine Activity Theory (RAT) Categorization examines the Motivated Offender (individual hacker, cybercriminal group, or AI attacker), the Suitable Target (victim type and data/assets targeted), and the Lack of Capable Guardian (effectiveness of security measures or lack thereof). The second tier AI-enhanced Social Engineering Parameters covers: Reconnaissance (traditional OSINT vs. AI-powered data mining), Deception (phishing, vishing, deepfakes), Attack Vectors (email, social media, AI-generated spear-phishing), Exploitation (attack success rates), and Crime (financial loss, reputational damage, regulatory consequences).

A comparative analysis was conducted to identify differences between traditional and AI-driven social engineering attacks, with a particular focus on their targeted vulnerabilities, success rates, and the role of automation in scaling attacks. The initial search yielded 183 papers, which were narrowed to 27 through thematic coding.

Results

The comparative analysis of cases revealed major shifts in how social engineering attacks are executed when AI technologies are involved. The following table summarizes the core differences observed between traditional and AI-enhanced attacks:

Table 1. A Comparison Between Traditional and AI-enhanced Social Engineering

Aspect	Traditional Social Engineering	AI-enhanced Social Engineering
Reconnaissance	Manual OSINT research	Automated, AI-driven data mining
Deception	Manually written phishing emails	AI-generated emails, deepfake calls/videos
Attack Vectors	Emails, phone calls	Emails, phone calls, chatbots, social media
Exploitation	Limited by human effort	Adapts in real time to user responses

Increased Efficiency in Reconnaissance: AI significantly enhances the data-gathering process through automated OSINT tools, facial recognition technology, AI-driven threat intelligence platforms, and deepfake and misinformation campaigns, allowing for faster and more precise targeting (Webasha, 2025). Lakshmanan (2025) found that over 57 threat actors with ties to China, Iran, North Korea, and Russia are using AI technology to further enable their information operations, gaining efficiency in productivity. AI can increase OSINT and reconnaissance tenfold, scouring the internet for any possible matches and even finding where they appear online by uploading a photo of the victim (The Hacker News, 2025).

Higher Success Rates in Deception: Deepfake technology and AI-generated phishing emails demonstrated a higher likelihood of deceiving victims compared to traditional social engineering techniques. According to Concannon (2024), AI-driven phishing attacks have risen by 50% over the past year. Furthermore, the FBI (2024) warned of a rise in AI-assisted phishing attacks, crafting highly personalized and convincing emails that make detection harder. A finance worker in Hong Kong was tricked into paying $25 million by fraudsters using a video call with whom he thought were all members of staff, but were deepfake recreations (Chen & Magramo, 2024). The deepfake recreations were so lifelike, he set aside his doubts he had about this video call after attending. Deepfake phishing has seen a 3,000% increase in 2023 and has been fueled by the accessibility and advancements of AI (Olney, 2024), and about every 5 minutes is when a deepfake attempt occurs (Onfido, 2025). In the first quarter of 2023, deepfake incidents increased by 245% year-over-year worldwide, with a 303% increase in the U.S. (Proofpoint). Social media has become increasingly used as a news source, with an 11% increase, with this information including over 1,150 unreliable AI-generated news websites (The Hacker News, 2025).

Diversified Attack Vectors: Traditional social engineering relies heavily on email and phone-based phishing, whereas AI expands methods to include chatbots, social media manipulation, and deepfake audio (Lakshmanan, 2025). Deepfake technology has advanced to the point where video and audio manipulation are nearly indistinguishable from real recordings. Cybercriminals are leveraging AI-generated deepfake calls to impersonate executives, government officials, and family members, leading to significant security and financial breaches. Giuffrida (2025) documents a case where AI-generated deepfake audio scams targeted Italian business leaders, including Giorgio Armani, resulting in financial losses exceeding millions of dollars. Microsoft has even developed an AI program that can clone your voice from only a 3-second audio clip, making it easier for cybercriminals to scam and commit identity fraud (Kan, 2023). A new AI-assisted ransomware, FunkSec, uses double extortion tactics to pressure victims into paying ransoms, with associated cybercriminals showing unsettling convergence of tactics, techniques, and objectives (Lakshmanan, 2025). A new AI-enhanced scam is also using social media ads that lead to phishing websites,

harvesting personal data once clicked, then using that information to directly contact and manipulate victims (Lakshmanan, 2024). AI-driven scams were also used on LinkedIn, which was estimated to have stolen more than $10 million worth of cryptocurrency in their social engineering campaign lasting six months (Lakshmanan, 2024). The adoption of AI-generated exploit code poses a significant challenge, too, especially considering the strategic use of zero-day vulnerabilities exploited by threat actors, achieving significant disruptions (NTTData, 2024). In 2023, phishing attacks were one of the top vectors in cybercrime, with the average breach costing around $4.76 million, and the use of generative AI making these attacks even easier to bypass defenses (Samala, 2024).

Greater Exploitation of Psychological Vulnerabilities: AI-generated content adapts in real time, tailoring messages to victims' responses, making attacks more convincing (Schmitt & Flechais, 2023). AI's ability to analyze large datasets enables attackers to identify high-value targets more efficiently. Attackers can use AI to monitor social media, scrape personal information, and generate detailed psychological profiles, ensuring that attacks are hyper-personalized. This level of precision increases the success rate of phishing and deception campaigns.

Automation and Scalability of Attacks: AI-driven attacks can be executed on a much larger scale than traditional social engineering. Automated AI models can generate and distribute phishing emails, create realistic deepfakes, and simulate human-like conversations across multiple platforms simultaneously. This scalability makes AI-assisted attacks more difficult to mitigate and respond to in real time (Stanham, 2025). Abnormal surveyed 300 cybersecurity leaders and found nearly 50% confirmed the presence of AI-powered attacks in their email environments, and this number is only set to increase. Amos (2023) studied FraudGPT, used to create bank-related phishing emails, where users only need to format their questions to include the bank's name, and FraudGPT will suggest where to implement suspicious links that lead to AI-created scam landing pages that ask for victim information.

Reduction in Attack Costs: Traditionally, social engineering attacks required extensive manual effort to research and craft convincing narratives. AI now enables attackers to automate these processes, significantly lowering the barrier to entry for cybercriminals. AI makes cybercrime more accessible due to the low risk and cost compared to traditional offenses (Manky & Baram, 2025). Heiding, Schneier, & Vishwanath (2024) in their recent study found that entire phishing processes can be automated using LLMs, reducing the costs of these attacks by more than 95%. The Hacker News (2023) found that 99% of those using ChatGPT claimed some form of cost-savings, and 25% said it reduced savings by $75,000 or more. Voice cloning companies for creating deepfakes or vishing attacks, allowing anyone to upload someone's voice for a monthly subscription of only $5 (Verma & Oremus, 2023).

Challenges in Detection and Defense: AI-driven attacks evolve rapidly, making traditional detection methods obsolete. Signature-based security systems struggle to keep up with dynamically generated AI content. The Mimecast Global Threat Intelligence Report (2024) emphasizes the need for AI-powered detection tools that can identify AI-generated manipulation attempts in real time. The Department of Justice recently seized two internet domains and searched nearly 1,000 social media accounts in which Russian threat actors were spreading disinformation (Lakshmanan, 2024). These bots were able to avoid bans and detection through X by blending into the social media environment. The 2024 Microsoft Digital Defense Report warns of sophisticated AI-enabled human targeting, where threats will be even more difficult to detect, even with AI tools assisting in defensive strategies.

These findings highlight AI's transformative impact on social engineering and emphasize the urgent need for adaptive cybersecurity measures. The evolution of AI-driven cybercrime presents new challenges for

law enforcement, regulatory agencies, and private sector organizations in maintaining security and trust in digital communications.

Discussion

Insights from the coded data informed the creation of Figure 1, a visual model representing the AI-driven social engineering attack cycle. The figure integrates RAT with the derived parameters to map a sequential attack process, showing how AI tools enhance each phase. This schematic guided interpretation of how AI shifts the social engineering threat landscape, how emerging technologies amplify the capabilities of attackers, and exposes critical vulnerabilities in digital communications and organizational defense mechanisms.

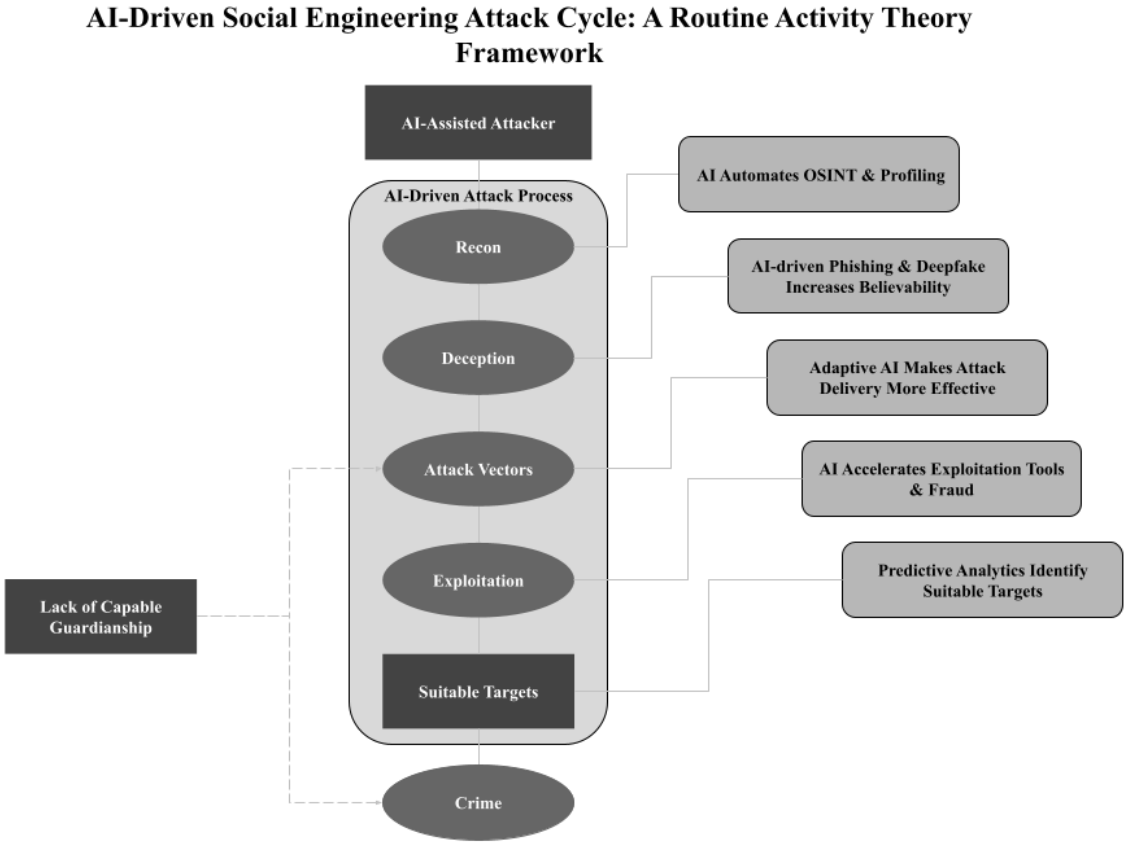

Figure 1. AI-driven social engineering attack cycle and enhancements made

This research reveals how AI redefines social engineering attacks by increasing automation, deception, precision, and scalability, creating an urgent need for adaptive cybersecurity strategies grounded in criminological theory. AI isn't just another tool in a hacker's kit—it fundamentally changes the speed, scale, and success of social engineering attacks. From corporate fraud to identity theft, these threats affect public trust, data integrity, and the very structure of digital society and security. By applying Routine Activity Theory, we provide a criminological framework to understand and disrupt these emerging attack patterns. The fusion of RAT with technical parameters provides interdisciplinary insights into: How AI lowers the barrier for attackers; Why current defense models (firewalls, training, endpoint protection) may

not detect AI-generated threats; Which phases of the social engineering cycle are most vulnerable to AI enhancement?

Implications

The emergence of AI-enhanced social engineering attacks signals a significant shift in the cybersecurity threat landscape. This study's findings reveal that attackers now wield tools that adapt in real time, scale effortlessly, and exploit psychological and behavioral vulnerabilities with unprecedented accuracy. These developments carry significant implications across three primary domains: policy, cybersecurity practice, and law enforcement. This paper makes an important contribution to future research by systematically reviewing the combination of AI, social engineering, and Routine Activity Theory. It provides valuable insights into where RAT applies, where it's inapplicable, and where there is room for improvement. By highlighting the challenges AI brings to social engineering and the limitations of the current framework, this paper aims to inform criminological theory, legal, and policy efforts to address the realities of AI-facilitated crime. Without this systematic examination, scholars and researchers will lack the critical knowledge to improve legal frameworks with the evolving forms of cybersecurity threats today.

As AI becomes increasingly integrated into both offensive and defensive cyber strategies, regulatory bodies must urgently address the gray areas surrounding AI use. First, there is a pressing need for comprehensive guidelines on the ethical and secure deployment of AI technologies in cybersecurity. These guidelines should define responsible AI development, especially regarding generative models that are capable of producing convincing phishing emails or deepfake content. Additionally, stricter incident reporting standards must be enforced across all sectors. Organizations should be required to disclose the use of AI in both cyberattacks and defense to improve collective threat intelligence. International cooperation is also essential, given the borderless nature of cybercrime. A coordinated framework could help standardize definitions, threat categorizations, and cross-jurisdictional prosecution procedures.

The current industry standard—user education and awareness training—remains critical but is no longer sufficient by itself. AI-enhanced attacks can bypass even well-trained employees by exploiting deepfake technology and natural language generation. Therefore, organizations must adopt a layered defense strategy incorporating AI-enabled tools capable of real-time threat detection and automated incident response. These tools should be trained on the same types of data used by attackers, allowing for proactive defense. Additionally, integrating the Routine Activity Theory (RAT) into cybersecurity risk assessments could provide a structured lens for identifying vulnerable systems and gaps in defenses (capable guardians).

From a legal and criminological standpoint, AI introduces unique challenges to attribution and prosecution. Existing cybercrime laws are often ill-equipped to handle crime facilitated by autonomous or semi-autonomous AI systems. For example, should deepfake voice scams be classified as fraud, impersonation, or something else entirely? Law enforcement agencies must evolve their investigative frameworks to include digital forensic techniques capable of detecting AI-generated content. Furthermore, legal definitions need to be expanded to encompass AI-assisted cybercrime, allowing courts to distinguish between human-led attacks and those augmented or initiated by AI. Without such evolution, there is a risk of legal loopholes that weaken accountability and justice.

Limitations

This study provides a framework for analyzing AI-enhanced social engineering through the RAT lens, but limitations should be noted. Data was gathered from reputable sources such as cybersecurity databases,

incident reports, and news articles, though it excluded some sources, particularly in private firms. This analysis focused on incidents from the past five years to ensure relevance, but this may have missed emerging or less-documented attack techniques.

Future Research Directions

The dynamic nature of AI-enhanced cyber threats calls for ongoing research into the convergence of AI, cybersecurity, and criminology. Future research should focus on developing RAT-informed AI models to detect social engineering in real-time, analyzing anomalies, sentiment, message tones, and behavioral cues to flag suspicious behavior. Additionally, criminological theories like RAT will need revision to account for both AI as an attacker and as a tool. Legal research should address intent, responsibility, and liability when AI acts autonomously, as well as explore how criminal justice systems conceive harm and punishment in AI-mediated offenses.

Conclusion

The integration of AI into social engineering tactics shows a pivotal shift in the cybersecurity and cybercrime landscape, challenging legal frameworks such as Routine Activity Theory. RAT remains a useful lens to examine criminal activity, but the rise of AI requires a reexamination of its core assumptions, mainly regarding the capabilities of offenders and the role of guardianship in digital environments. The comparative analysis of traditional and AI-enhanced social engineering attacks highlights the scale, speed, and sophistication of these attacks, with which AI is reshaping cyber threats. These enhancements suggest that unchanging criminological models will fall short when addressing the complexities introduced by emerging AI technologies. Therefore, Routine Activity Theory, in its current form, does not adequately address the pressing issues of AI's evolution in social engineering crimes.

To effectively combat AI-driven social engineering attacks, security professionals, policy makers, researchers, and law enforcement must collaborate in evolving theoretical models and defense mechanisms. Future adaptations of RAT must account for the ever-changing, adaptive nature of AI and its impact on cybercriminal reach, victim behavior, and the capabilities of digital guardianship. As cyber threats continue to grow in complexity, interdisciplinary approaches that merge artificial intelligence, cybersecurity, and criminology will be crucial in creating informed, adaptive policies and resilient defenses in the new digital age.

References

Abnormal. *How AI-Enabled Cyberattacks Work, Why They're Increasing, and How to Stop Them*. https://abnormalsecurity.com/glossary/ai-enabled-cyberattacks

Ahmad, R., & Thurasamy, R. (2022). A Systematic Literature Review of Routine Activity Theory's Applicability in Cybercrimes. Journal of Cyber Security and Mobility. https://doi.org/10.13052/jcsm2245-1439.1133.

Alahmed, Y., Abadla, R., & Ansari, M. (2024). Exploring the Potential Implications of AI-generated Content in Social Engineering Attacks. 2024 International Conference on Multimedia Computing, Networking and Applications (MCNA), 64-73. https://doi.org/10.1109/MCNA63144.2024.10703950.

Amos, Z. (2023, August 11). *What Is FraudGPT?* HackerNoon. https://hackernoon.com/what-is-fraudgpt

Blauth, T., Gstrein, O., & Zwitter, A. (2022). Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. IEEE Access, 10, 77110-77122. https://doi.org/10.1109/access.2022.3191790.

Blessing, G., Azeta, A., Misra, S., Osamor, V., Sanz, L., & Pospelova, V. (2022). The Emerging Threat of Ai-driven Cyber Attacks: A Review. Applied Artificial Intelligence, 36. https://doi.org/10.1080/08839514.2022.2037254.

Broadhurst, R., Maxim, D., Brown, P., Trivedi, H., & Wang, J. (2019). Artificial Intelligence and Crime. Types of Offending eJournal. https://doi.org/10.2139/ssrn.3407779.

Caldwell, M., Andrews, J., Tanay, T., & Griffin, L. (2020). AI-enabled future crime. Crime Science, 9. https://doi.org/10.1186/s40163-020-00123-8.

Chen, H., & Magramo, K. (2024, February 4). *Finance worker pays out $25 million after video call with deepfake 'chief financial officer'*. CNN. https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

Concannon, M. (2024, July 16). *AI in Social Engineering: The Next Generation of Cyber Threats*. Ntiva. https://www.ntiva.com/blog/ai-social-engineering-attacks?utm_source=chatgpt.com

Giuffrida, A.(2025, February 10). *AI phone scam targets Italian business leaders including Giorgio Armani*. The Guardian. https://www.theguardian.com/world/2025/feb/10/ai-phone-scam-targets-italian-business-leaders-including-giorgio-armani?utm_source=chatgpt.com

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. IEEE Access, 11, 80218-80245. https://doi.org/10.1109/ACCESS.2023.3300381.

Hatfield, J. (2018). Social engineering in cybersecurity: The evolution of a concept. Comput. Secur., 73, 102-113. https://doi.org/10.1016/J.COSE.2017.10.008.

Hayward, K., & Maas, M. (2020). Artificial intelligence and crime: A primer for criminologists. Crime, Media, Culture: An International Journal, 17, 209 - 233. https://doi.org/10.1177/1741659020917434.

Heiding, F., Schneier, B., & Vishwanath, A. (2024, May 30). *AI Will Increase the Quantity — and Quality — of Phishing Scams*. HBR. https://hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams

Holt, T., & Bossler, A. (2008). Examining the Applicability of Lifestyle-Routine Activities Theory for Cybercrime Victimization. Deviant Behavior, 30, 1 - 25. https://doi.org/10.1080/01639620701876577.

Jeong, D. (2020). Artificial Intelligence Security Threat, Crime, and Forensics: Taxonomy and Open Issues. IEEE Access, 8, 184560-184574. https://doi.org/10.1109/ACCESS.2020.3029280.

Kan, M. (2023, January 10). *Microsoft's AI Program Can Clone Your Voice From a 3-Second Audio Clip*. PCMag. https://www.pcmag.com/news/microsofts-ai-program-can-clone-your-voice-from-a-3-second-audio-clip

Kayser, C., Mastrorilli, M., & Cadigan, R. (2019). Preventing cybercrime: A framework for understanding the role of human vulnerabilities. Cyber Security: A Peer-Reviewed Journal. https://doi.org/10.69554/sfox9866.

Lakshmanan, R. (2025, January 31). *Top 5 AI-Powered Social Engineering Attacks*. The Hacker News. https://thehackernews.com/2025/01/top-5-ai-powered-social-engineering.html?utm_source=chatgpt.com

Lakshmanan, R. (2025, January 30). *Google: Over 57 Nation-State Threat Groups Using AI for Cyber Operations*. The Hacker News. https://thehackernews.com/2025/01/google-over-57-nation-state-threat.html

Lakshmanan, R. (2025, January 10). *AI-Driven Ransomware FunkSec Targets 85 Victims Using Double Extortion Tactics*. The Hacker News. https://thehackernews.com/2025/01/ai-driven-ransomware-funksec-targets-85.html?m=1

Lakshmanan, R. (2024, December 16). *New Investment Scam Leverages AI, Social Media Ads to Target Victims Worldwide*. The Hacker News. https://thehackernews.com/2024/12/new-investment-scam-leverages-ai-social.html

Lakshmanan, R. (2024, November 23). *North Korean Hackers Steal $10M with AI-Driven Scams and Malware on LinkedIn*. The Hacker News. https://thehackernews.com/2024/11/north-korean-hackers-steal-10m-with-ai.html

Lakshmanan, R. (2024, July 12). *U.S. Seizes Domains Used by AI-Powered Russian Bot Farm for Disinformation*. The Hacker News. https://thehackernews.com/2024/07/us-seizes-domains-used-by-ai-powered.html?m=1

Leukfeldt, E., & Yar, M. (2016). Applying Routine Activity Theory to Cybercrime: A Theoretical and Empirical Analysis. Deviant Behavior, 37, 263 - 280. https://doi.org/10.1080/01639625.2015.1012409.

Manky, D., & Baram, G. (2025). *Beyond Phishing: Exploring the Rise of AI-enabled Cybercrime*. CLTC. https://cltc.berkeley.edu/2025/01/16/beyond-phishing-exploring-the-rise-of-ai-enabled-cybercrime/#:~:text=Lower%20Barriers%2C%20Wider%20Participation&text=By%20lowering%20the%20technical%20barrier,to%20traditional%20street%2Dlevel%20offenses.

Microsoft. (2024). *Microsoft Digital Defense Report 2024*. https://www.microsoft.com/en-us/security/security-insider/intelligence-reports/microsoft-digital-defense-report-2024

Mimecast. (2024). *Global Threat Intelligence Report 2024 H1*. https://www.mimecast.com/resources/ebooks/global-threat-intelligence-report-january-june-2024/

Mohsin, K. (2020). Regulation of AI and AI Crimes. . https://doi.org/10.2139/ssrn.3552140.

NTTData. (2024). *Global Threat Intelligence Report*. https://us.nttdata.com/en/insights/global-threat-intelligence-report

Olney, M. (2024, March 5). *How is AI changing Social Engineering attacks?* Integrity360. https://insights.integrity360.com/how-is-ai-changing-social-engineering-attacks

Onfido. (2025). *2025 Identity Fraud Report*. https://onfido.com/landing/identity-fraud-report/

Proofpoint. *Deepfake Technology*. https://www.proofpoint.com/us/threat-reference/deepfake

Samala, M. (2024, August 5). *The Evolution Of Social Engineering And Phishing In The Age Of Artificial Intelligence*. Lumen. https://blog.lumen.com/the-evolution-of-social-engineering-and-phishing-in-the-age-of-artificial-intelligence/

Schmitt, M., & Flechais, I. (2023). Digital Deception: Generative Artificial Intelligence in Social Engineering and Phishing. Artif. Intell. Rev., 57, 324. https://doi.org/10.48550/arXiv.2310.13715.

Stanham, L. (2025, January 16). *AI-Powered Cyberattacks*. CrowdStrike. https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/ai-powered-cyberattacks/?utm_source=chatgpt.com

The Hacker News. (2025, February 21). *AI-Powered Deception is a Menace to Our Societies*. https://thehackernews.com/2025/02/ai-powered-deception-is-menace-to-our.html

The Hacker News. (2025, February 14). *AI-Powered Social Engineering: Ancillary Tools and Techniques*. https://thehackernews.com/2025/02/ai-powered-social-engineering-ancillary.html

The Hacker News. (2023, June 26). *How Generative AI Can Dupe SaaS Authentication Protocols — And Effective Ways To Prevent Other Key AI Risks in SaaS*. https://thehackernews.com/2023/06/how-generative-ai-can-dupe-saas.html?m=1

Verma, P., & Oremus, W. (2023, October 14). *AI voice clones mimic politicians and celebrities, reshaping reality*. The Washington Post. https://www.unionleader.com/ai-voice-clones-mimic-politicians-and-celebrities-reshaping-reality/article_f3c03b52-6d5f-56bc-aa39-05681c344a21.html

Webashsa Technologies. (2025, February 24). *The Rise of AI in Reconnaissance and Data Gathering | How Artificial Intelligence is Transforming Intelligence Collection, Cybersecurity, and Surveillance*. https://www.webasha.com/blog/the-rise-of-ai-in-reconnaissance-and-data-gathering-how-artificial-intelligence-is-transforming-intelligence-collection-cybersecurity-and-surveillance