Ein Überblick über Social Engeneering und wie Künstliche Intelligenz darauf einwirkt

Max Becker
Hochschule Rhein Main
Wiesbaden, Deutschland

***Abstract*—Social Engineering stellt eine der bedeutendsten Bedrohungen für die Cybersicherheit dar, da es gezielt menschliche Schwachstellen wie Vertrauen, Hilfsbereitschaft und Angst ausnutzt. Mit der zunehmenden Verbreitung Künstlicher Intelligenz (KI) hat sich das Spektrum solcher Angriffe erweitert. KI-Technologien ermöglichen es Angreifern, personalisierte Angriffe mit größerer Präzision und Effizienz durchzuführen, während sie gleichzeitig neue Verteidigungsstrategien unterstützen. Dieses Paper fasst den generellen Ablauf von Social Engineering Angriffen und deren Klassifikationen zusammen, ergänzt durch eine Analyse der Rolle von KI. Dabei wird aufgezeigt, wie KI-basierte Methoden, wie Natural Language Processing und maschinelles Lernen, sowohl die Erstellung realistischer Angriffe als auch deren frühzeitige Erkennung und Abwehr revolutionieren.**

I. Einleitung

In der heutigen digitalen Welt stellt Social Engineering eine erhebliche Bedrohung für die Cybersicherheit dar. Diese Technik nutzt menschliche Schwächen wie Vertrauen und Hilfsbereitschaft aus, um unbefugten Zugang zu sensiblen Informationen zu erlangen. Trotz fortschrittlicher technischer Sicherheitsmaßnahmen bleibt der Mensch ein verwundbares Glied in der Sicherheitskette. [1]

Mit der rasanten Entwicklung der Künstlichen Intelligenz (KI) hat sich die Dynamik von Social-Engineering-Angriffen verändert. KI-Technologien ermöglichen es Angreifern, große Datenmengen zu analysieren und menschenähnliche Interaktionen zu simulieren, was die Effektivität und Präzision solcher Angriffe erhöht [6]. Gleichzeitig bietet KI auch innovative Verteidigungsmechanismen, die helfen, Bedrohungen frühzeitig zu erkennen und abzuwehren [7].

Die Bedrohung durch Social Engineering ist im Themenbereich der Cybersicherheit allgegenwärtig, das ergibt sich aus der Unvermeidbarkeit menschlicher Schwachstellen in der Cybersicherheit. Es gibt kein Computersystem auf der Welt, das nicht von Menschen abhängt, ganz gleich, wie gut die Sicherheitsmaßnahmen konzipiert und umgesetzt sind. Diese beteiligten menschlichen Elemente sind deutlich verwundbarer als viele andere Sicherheitsschwachstellen [2]. Das bedeutet, dass diese Sicherheitsschwäche universell und unabhängig von Plattform, Software, Netzwerk oder Alter der Geräte ist [3]. Es ist möglich, ein Vermögen auszugeben, um Technologie und Dienstleistungen von jedem auf dem Markt vertretenen Anbieter zu kaufen, um sein System zu schützen. Anfällig für menschliche Manipulation bleibt es trotzdem. [4].

Eine Studie des Unternehmens Accenture stellte fest, dass im Jahr 2018 85% aller betrachteten Unternehmen von Social-Engineering-Angriffen betroffen waren, was einem Anstieg von 16% gegenüber dem Vorjahr entspricht. Die durchschnittlichen jährlichen Kosten von Social-Engineering-Angriffen für Unternehmen beliefen sich 2018 auf mehr als 1,4 Millionen US-Dollar, was einem Anstieg von 8% gegenüber dem Vorjahr entspricht [5].

Dieses Paper soll daher einen groben Überblick über die Methoden von Social-Engineering geben, als auch die duale Rolle der KI im Bereich der Cybersicherheit zusammenfassen. Da aus der Anwendung von KI im Kontext von Social Engineering sowohl Risiken als auch Chancen für die Cybersicherheit entstehen, ist es das Ziel, ein tieferes Verständnis für die Herausforderungen und Möglichkeiten zu schaffen, die KI in diesem Bereich bietet.

II. Hintergrund: Social Engineering und Künstliche Intelligenz

”Beim Social Engineering werden menschliche Eigenschaften wie Hilfsbereitschaft, Vertrauen, Angst oder Respekt vor Autorität ausgenutzt, um Personen geschickt zu manipulieren. Cyber-Kriminelle verleiten das Opfer auf diese Weise beispielsweise dazu, vertrauliche Informationen preiszugeben, Sicherheitsfunktionen auszuhebeln, Überweisungen zu tätigen oder Schadsoftware auf dem privaten Gerät oder einem Computer im Firmennetzwerk zu installieren.” So das BSI zur Frage ”Was ist Sozial Engineering?”. [1]

A. Genereller Ablauf von Social-Engineering Angriffen

Social-Engineering Angriffe nutzen eine Vielzahl Techniken, um an sensible Informationen des Opfers zu kommen. Unabhängig davon bleibt das Muster der Angriffe gleich. Der Angriffsprozess beinhaltet 4 Etappen. Diese sind das Sammeln von Daten und Informationen, Verbesserung der Beziehungen und Vertrauensbildung, Ausnutzung und schließlich Ausführung. Die erste Phase eines Social-Engineering-Angriffs umfasst Arbeit zur Sammlung von Informationen über die Zielperson. Dies ist die wichtigste Phase des Angriffs, da alle weiteren Phasen und das Endergebnis von den in dieser Phase gewonnenen Informationen abhängen. Nach der Nachforschung fokussiert sich der Angreifer darauf, bei der Zielperson Vertrauen zu gewinnen. Die Ausnutzungs-Phase besteht aus dem Manipulieren der Zielperson, um an Informationen, wie zum Beispiel Passwörter, zu gelangen

oder einen sicherheitskritischen Fehler zu provozieren. Zum Schluss, in der Ausführungsphase, werden alle möglichen Informationen und Ressourcen gesammelt. Zu guter Letzt versucht der Angreifer, seine Spuren zu verwischen. [10]

B. Klassifizierung von Sozial-Engineering-Angriffen

In Übereinstimmung mit der Klassifizierung von Hussain Aldawood und Geoffrey Skinner [10], dargestellt in 1, können Sozial-Engineering-Angriffe anhand ihrer Typen, Anwender und Medium des Angriffs kategorisiert werden.

1) Typen von Social Engineering Angriffen: Social-Engineering-Angriffe sind vielschichtig und umfassen mehrere soziale, technische und physische Aspekte. Diese Aspekte werden normalerweise in verschiedenen Phasen des eigentlichen Angriffs eingesetzt. [10]

Der soziale Aspekt Social-Engineering basiert auf Manipulation und Überredung. Dabei werden sozialpsychologische Techniken wie die Überzeugungsprinzipien genutzt, um das Opfer auf eine Weise zu manipulieren, durch welche sensible Informationen freigegeben werden. Was unter sensible Informationen zu verstehen ist, wird in Kapitel 3.a aufgegriffen. [10], [11]

Technische Social-Engineering Angriffe werden meist über das Internet ausgeführt. Ungesicherte soziale Netzwerke sind besonders interessant, da sie Zugang zu persönlichen Informationen und möglichen Passwörtern ermöglichen. E-Mail Anhänge, Pop-up-Fenster und Webseiten gehören zu den technischen Werkzeugen von Sozial-Engineers. [10], [11]

Manche Vorgehensweisen von Social-Engineering benötigen eine physikalische Komponente. Wie der Name vermuten lässt, muss der Angreifer eine physische Aktion durchführen, um an Informationen über die Zielperson zu gelangen. Ein Beispiel hiervon ist eine Methode mit dem Name "Dumpster Diving". Dies beinhaltet das Durchsuchen von Entsorgtem einer Organisation, um die persönlichen Daten von Mitarbeitern, Handbücher, Notizen und sogar Ausdrucke sensibler Informationen, wie z. B. Benutzeranmeldedaten, finden zu können.

Die meisten Social-Engineering Angriffe fassen mehrere oder sogar alle vorher erwähnten Typen zusammen. Diesem Socio-technischen Ansatz entspringen die bekanntesten und gefährlichsten Angriffsarten, wie zum Beispiel Baiting oder Phishing. Ein sozialer Ansatz hilft beim Aufbau einer vertrauensvollen Beziehung, während der technische Ansatz eine Möglichkeit bietet, Zugang zu sensiblen persönlichen Informationen zu erhalten. Daher wird bei einem wirksamen soziotechnischen Angriff eine Kombination aus beidem verwendet. [10], [11]

2) In Person und Technologie gestützte Interaktionen: Angriffe wie zum Beispiel Tailgaiting werden als in Person Interaction eingegliedert. Bei dieser Art von Social-Engineering-Angriff schleicht sich ein Angreifer einfach hinter einer Person ein, die legitimen Zugang zu einem bestimmten eingeschränkten Bereich hat. Der Höflichkeit halber hält die authentische Person dem Angreifer in der Regel die Tür auf. Die entscheidende Eigenschaft von in Person Social-Engineering Angriffen ist demnach, wie der Name vermuten lässt, dass sie auf direkter menschlicher Interaktion basieren.

Technologie gestützte Social-Engineering-Angriffe nutzen digitale Medien und das Internet, um Opfer zu täuschen und an Informationen zu gelangen. Dabei kommen Geräte wie Computer, Smartphones oder Tablets zum Einsatz. Typische Methoden sind Phishing-E-Mails, gefälschte Websites, Schadsoftware, Manipulation in sozialen Netzwerken, technische Täuschungen wie Spoofing sowie Angriffe über mobile Apps oder Messaging-Dienste. [10]

C. Ziele von Social Engineering Angriffen

Individuelle Informationen gehören zu den häufigsten Zielen von Sozial-Engineering-Angriffen. Hierbei geht es um persönliche Daten wie Sozialversicherungsnummern, Geburtsdaten oder Passwörter. Diese Informationen können für Identitätsdiebstahl oder betrügerische Aktivitäten verwendet werden. Finanzdaten wie Kreditkartennummern und Bankverbindungen sind ebenfalls hochgradig attraktiv, da sie Angreifern direkten Zugriff auf Vermögenswerte ermöglichen. Eine weitere Kategorie stellen Login-Daten dar. Ob für E-Mail-Konten, soziale Netzwerke oder geschäftliche Accounts – Zugangsdaten bieten oft die Möglichkeit, noch mehr private oder berufliche Informationen zu kompromittieren und ermöglichen Folgeangriffe. [11]

Auf Unternehmensebene stehen andere schützenswerte Ziele im Fokus. Besonders interessant sind geheime Geschäftsdaten, darunter Forschungsergebnisse, strategische Pläne oder Patente, die den wirtschaftlichen Erfolg eines Unternehmens gefährden können. Angreifer versuchen auch, Zugang zu internen Systemen zu erlangen. Netzwerke, Datenbanken und Cloud-Dienste eines Unternehmens sind nicht nur für deren Betrieb essenziell, sondern enthalten oft auch Kundendaten. Solche Daten sind nicht nur für Wettbewerber, sondern auch für Kriminelle von Interesse, die daraus Profit schlagen oder Erpressungen durchführen könnten. [11]

Kritische Infrastrukturen sind ein weiteres wichtiges Ziel. Technologische Systeme wie Energie-, Verkehrs- oder Kommunikationssysteme sind essenziell für den reibungslosen Ablauf moderner Gesellschaften. Angriffe auf diese Systeme, etwa durch das Umgehen von Sicherheitslösungen wie Token-Systemen, können verheerende Auswirkungen haben. Ein bekanntes Beispiel ist der RSA-Angriff von 2011, der bewies, wie effektiv solche Angriffe sein können, wenn Sicherheitsmaßnahmen ausgehebelt werden. [11]

Ein besonderer Fokus von Sozial-Engineering-Angriffen kann auf politisch oder wirtschaftlich motivierten Zielen liegen. Hierbei werden beispielsweise diplomatische oder Regierungsinstitutionen angegriffen. Ein Beispiel ist die "Red October"-Kampagne, bei der sensible Daten aus diplomatischen und staatlichen Netzwerken gestohlen wurden. Bei weiterem Interesse an der "Red October"-Kampagne bietet sich [14] an. Auch journalistische Organisationen stehen im Fadenkreuz, da sie oft über vertrauliche Berichte verfügen. Der Angriff auf die New York Times im Jahr 2013 verdeutlichte,

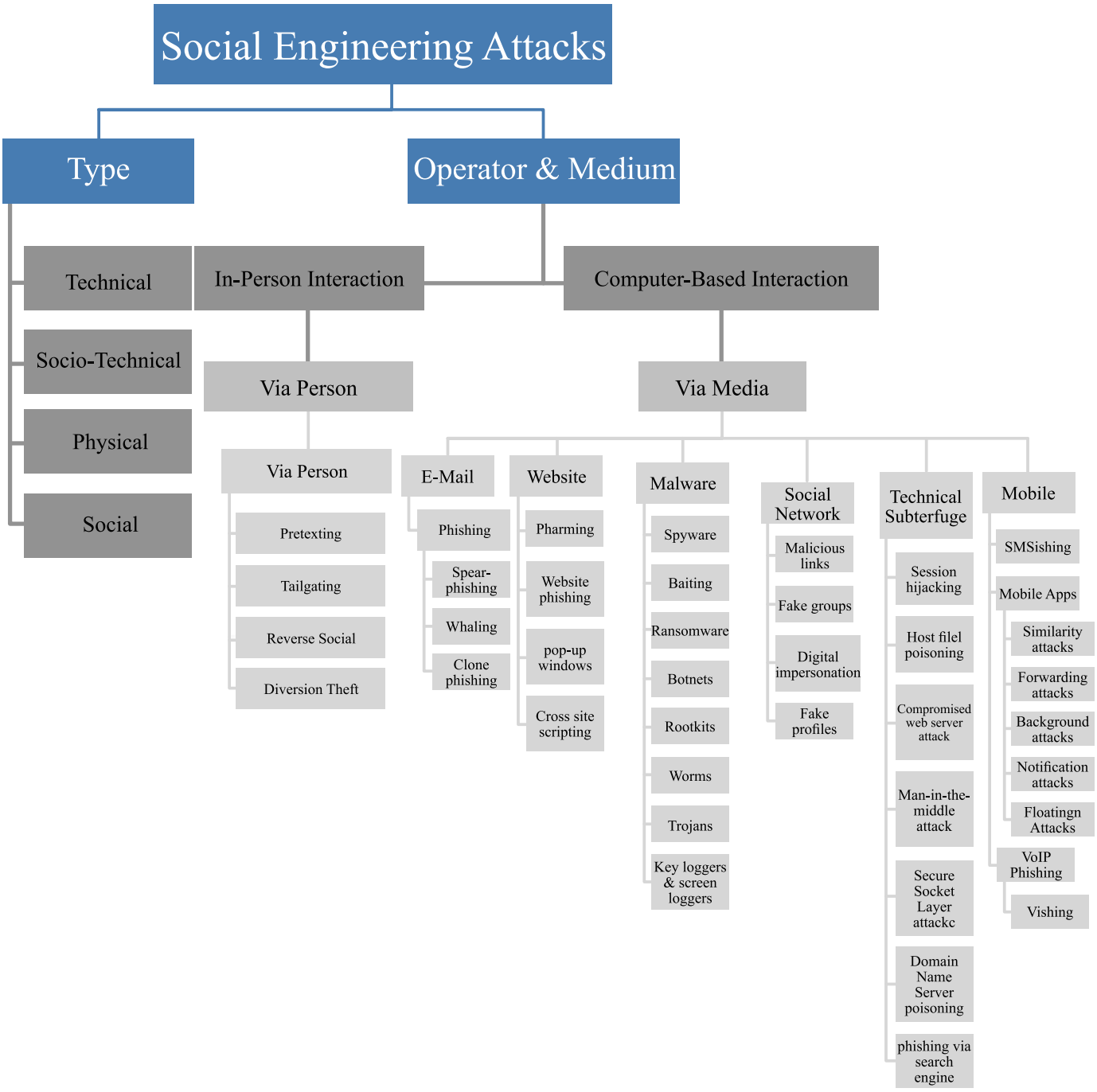

Fig. 1. Klassifizierung von Social Engineering Angriffen

wie wichtig die Absicherung solcher Organisationen ist, um die Unabhängigkeit des Journalismus zu gewährleisten. [11]

Nicht zuletzt sind auch öffentliche Netzwerke und Cloud-Dienste beliebte Ziele. Sensible Daten, die in der Cloud gespeichert sind, können leicht zu einer Goldgrube für Angreifer werden, wenn keine ausreichenden Sicherheitsmaßnahmen getroffen werden. Soziale Netzwerke spielen ebenfalls eine zentrale Rolle, da sie oft eine Fülle von Informationen über Einzelpersonen oder Unternehmen preisgeben. Diese Informationen können in der Folge für weitere Angriffe genutzt werden, sei es auf persönlicher oder beruflicher Ebene. [11]

D. Ablauf von Social Engineering Angriffen

Social Engineering Angriffe folgen einer systematischen Struktur, die sich in mehrere klar definierte Phasen gliedert. Jede Phase spielt eine entscheidende Rolle für den Erfolg eines Angriffs, da sie aufeinander aufbauen und sorgfältig geplant werden müssen. Kevin Mitnick formulierte in seinem Buch “The art of deception: controlling the human element

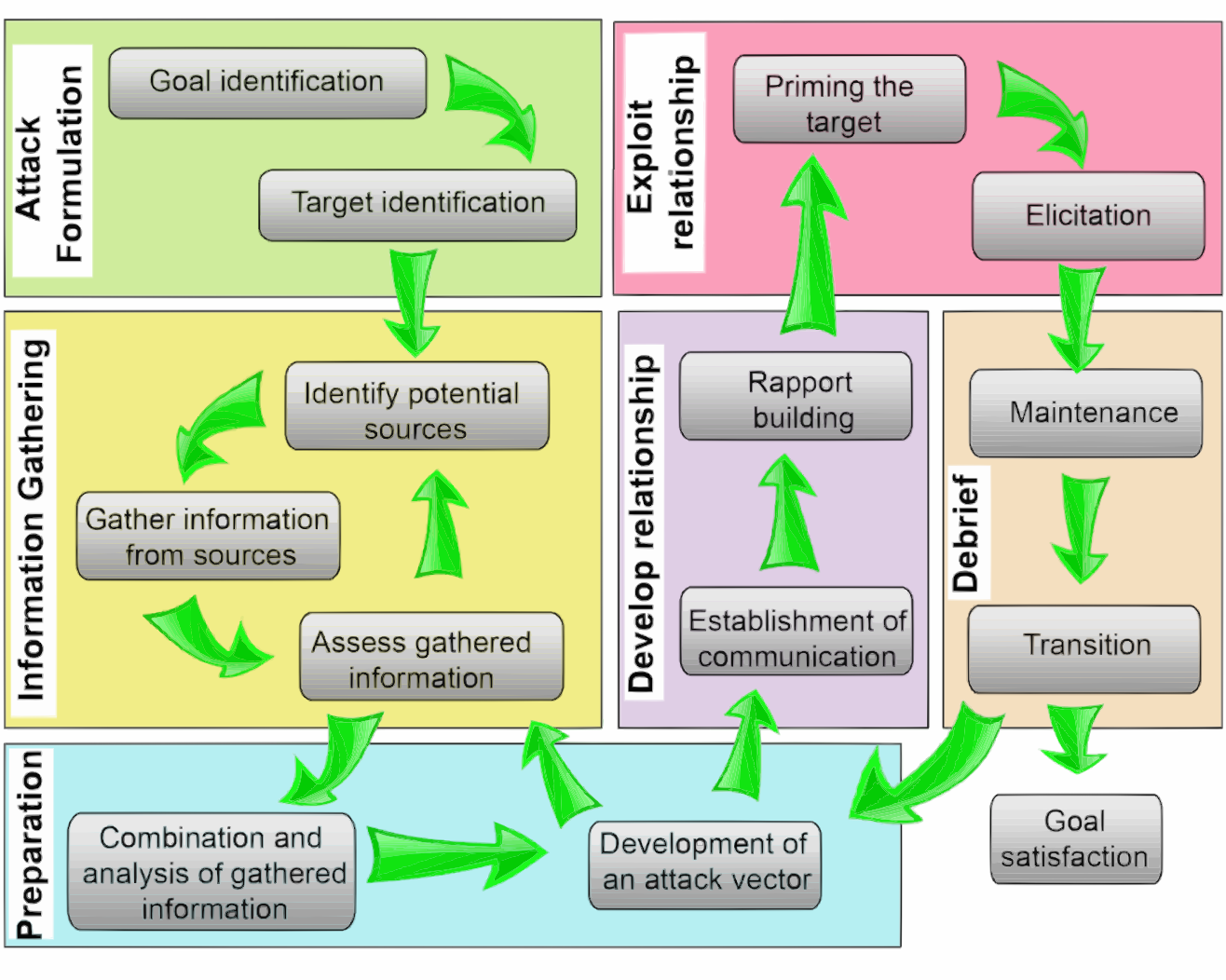

Fig. 2. Social Engineering Attack Framework [15]

of security“ ein Model, welches den Ablauf von Social Engineering Angriffen umrahmt. Dieses Model wurde später durch neue Schritte erweitert, um zeitliche und prozessuale Abläufe abzubilden. Das gesamte Framework ist dargestellt in 2. [13], [15]

1) Formulierung des Angriffs: Der Angriffsprozess beginnt mit der Definition des Ziels. Angreifer bestimmen zunächst, welche Informationen oder Handlungen sie durch den Angriff erzielen möchten. Daraufhin wird ein Ziel ausgewählt, das entweder eine Einzelperson oder eine Organisation sein kann. Beispiele sind IT-Mitarbeiter mit privilegierten Zugriffsrechten oder Angestellte im Kundensupport. [15]

2) Informationssammlung: In dieser Phase sammeln Angreifer möglichst viele Informationen über das Ziel. Quellen wie soziale Netzwerke, Unternehmenswebsites oder sogar unkonventionelle Methoden wie das Durchsuchen von Müll (Dumpster Diving) werden genutzt. Die gesammelten Daten werden anschließend bewertet, um sicherzustellen, dass sie für die weiteren Schritte ausreichen. Falls nötig, wird die Informationssammlung wiederholt. [15]

3) Vorbereitung: Mit den gesammelten Daten wird ein glaubwürdiges Angriffsszenario entwickelt. Diese Phase umfasst die Erstellung eines sogenannten Pretexting-Szenarios, also eines Vorwands, um das Vertrauen des Ziels zu gewinnen. Hierbei wird ein Angriffsmittel (z. B. eine Phishing-E-Mail) und das Kommunikationsmedium definiert, das in der nächsten Phase genutzt wird. [15]

4) Beziehungsaufbau: Das Ziel dieser Phase ist es, eine vertrauensvolle Beziehung zum Ziel aufzubauen. Der Angreifer nimmt Kontakt auf und nutzt zuvor gesammelte Informationen, um Authentizität zu simulieren. Beispiele sind das Zitieren interner Informationen oder das Vortäuschen einer Autoritätsposition. Der Erfolg dieser Phase hängt maßgeblich von der Glaubwürdigkeit des Vorwands ab. [15]

5) Ausnutzen der Beziehung: Nachdem eine Beziehung aufgebaut wurde, nutzt der Angreifer diese aus, um das Ziel zur Preisgabe von Informationen zu bewegen. Hierbei werden oft emotionale Manipulationstechniken eingesetzt, wie das

Schaffen von Dringlichkeit oder das Ausnutzen von Sympathie. Sobald der Angreifer die gewünschten Informationen erlangt hat, endet diese Phase. [15]

6) Debriefing und Übergang: Ein einzigartiger Aspekt des erweiterten Frameworks ist das Debriefing. Hierbei wird das Ziel wieder in einen neutralen emotionalen Zustand versetzt, um Verdachtsmomente zu vermeiden. Ist das Angriffsziel erreicht, wird der Angriff beendet. Falls zusätzliche Informationen benötigt werden, kehrt der Angreifer zu einer früheren Phase zurück, um den Prozess zu wiederholen. [15]

E. Grundlegende Begriffsklärung künstliche Intelligenz

Social Engineering Angreifer verwenden verschiedene Werkzeuge, um den Angriff für sie einfacher zu machen. Eine Gruppe solcher Werkzeuge kann man unter dem Begriff Künstliche Intelligenz (KI) zusammenfassen. Dadurch spielt KI eine zunehmend bedeutende Rolle im Bereich des Social Engineering, indem sie die Effektivität und Raffinesse solcher Angriffe erhöht. Eine der Hauptanwendungen von KI im Social Engineering ist die Analyse großer Datenmengen, die es Angreifern ermöglicht, gezielte und personalisierte Angriffe zu entwickeln [6], [12]. Maschinelles Lernen (ML), einem Teilbereich der Künstlichen Intelligenz, welches sich auf die Entwicklung von Algorithmen und Modellen, die Muster lernen und auf der Grundlage von Daten Vorhersagen oder Entscheidungen treffen können, konzentriert [12], kann zur Optimierung von Angriffstechniken genutzt werden [7].

Maschinelles Lernen kann in supervised learning, unsupervised learning, semisupervised learning und reinforcement learning unterteilt werden. Durchbrüche beim Deep Learning, einer Untergruppe des maschinellen Lernens, haben das Feld revolutioniert, indem sie das Training neuronaler Netze mit mehreren Schichten ermöglichten, was zu bemerkenswerten Fortschritten in Bereichen wie Bilderkennung, Verarbeitung natürlicher Sprache und autonomen Systemen führte. [12] Generative KI-Technologien, wie Deep Learning und in Erweiterung dazu Large Language Models (LLMs), ermöglichen demnach die Erstellung realistischer Inhalte, die menschliche Kommunikation nachahmen [8]. Diese Technologien werden genutzt, um überzeugende Phishing-Nachrichten oder Deepfake-Inhalte zu erstellen. Diese sind nur schwer von echten Interaktionen zu unterscheiden [9].

Natural Language Processing (NLP) ist ein besonders relevantes Teilgebiet generativer KI, denn es beschäftigt sich mit Sprachverarbeitung und -erkennung, was es Angreifern ermöglicht, automatisierte Systeme zu entwickeln, die auf natürliche Weise mit Opfern interagieren können [7].

Diese Fortschritte in der KI-Technologie stellen sowohl eine Herausforderung als auch eine Chance für die Cybersicherheit dar, da sie sowohl für Angriffe als auch für Verteidigungsmaßnahmen genutzt werden können.

III. Techniken KI-gestützter Angriffe

Generative KI umfasst ein breites Spektrum an fortschrittlichen Funktionen, die die Erstellung und Bearbeitung verschiedener Arten von Inhalten ermöglichen. Im Kontext von Social Engineering Angriffen lassen sich mehrere mögliche Funktionen herausfiltern.

A. KI-gestützte Social Engineering Angriffe

Generative Künstliche Intelligenz (KI) bietet Angreifern zahlreiche Möglichkeiten, die Effektivität von Social Engineering Angriffen zu steigern. Eine zentrale Rolle spielt hierbei die Fähigkeit von generativen KI-Algorithmen, realistische Inhalte wie Texte, Bilder, Videos oder Stimmen zu erzeugen. Diese können gezielt genutzt werden, um Angriffsvektoren wie Phishing-E-Mails oder Pharming-Angriffe zu erstellen. Besonders im Bereich des Online-Mediums ist diese Form der Täuschung schwer zu erkennen und ermöglicht Angreifern, Vertrauen bei den Opfern zu gewinnen und so sensible Informationen zu extrahieren.

Ein weiteres leistungsstarkes Werkzeug ist die KI-gestützte Datenanalyse. Mithilfe fortschrittlicher Algorithmen kann eine große Menge an Informationen effizient verarbeitet werden, um potenzielle Zielpersonen zu identifizieren und deren Schwachstellen zu bewerten. Basierend auf diesen Daten können Angreifer sogar das Verhalten ihrer Opfer vorhersagen und Angriffsmethoden gezielt anpassen. In Kombination mit KI-Scraping, einem automatisierten Verfahren zur Datenerfassung aus Quellen wie sozialen Netzwerken oder öffentlichen Datenbanken, entstehen detaillierte Profile, die eine präzise und personalisierte Ansprache ermöglichen.

Zudem unterstützt KI die Automatisierung redundanter Prozesse, wie die initiale Kontaktaufnahme mit Zielpersonen oder die Aufrechterhaltung einer konstanten Kommunikation. Diese Prozesse werden mit menschenähnlicher Präzision durchgeführt, was das Risiko für Angreifer minimiert. Besonders relevant sind in diesem Zusammenhang KI-gestützte Chatbots, die in der Lage sind, überzeugende Gespräche zu führen, Vertrauen aufzubauen und Informationen zu sammeln. Diese täuschend echten Interaktionen sind für Opfer oft kaum von menschlichen Gesprächen zu unterscheiden und stellen daher eine erhebliche Bedrohung dar.

Die Orchestrierung komplexer Angriffe wird durch KI ebenfalls erleichtert. Durch eine gezielte Koordination zwischen verschiedenen KI-Systemen bleibt die Kontinuität eines Angriffs erhalten, selbst wenn mehrere Mitwirkende beteiligt sind. Diese Fähigkeit ermöglicht es, groß angelegte Social Engineering Kampagnen einheitlich und effizient umzusetzen. Schließlich kann KI auch zur Bewertung und Optimierung von Angriffen eingesetzt werden. Analysen vergangener Angriffe liefern wichtige Erkenntnisse, um Strategien anzupassen und die Erfolgsaussichten zukünftiger Angriffe zu erhöhen. [12]

B. Defensive Anwendungen der KI gegen Social Engineering

Neben ihrem Einsatz für Angriffe, spielt KI auch eine entscheidende Rolle in der Verteidigung gegen Social Engineering Bedrohungen. Eine zentrale defensive Anwendung ist die KI-gestützte Bedrohungserkennung, die durch maschinelles Lernen ungewöhnliche Aktivitäten identifiziert. Anders als traditionelle signaturbasierte Systeme erkennt KI

auch bisher unbekannte Angriffsformen, indem sie Abweichungen von normalen Verhaltensmustern analysiert, statt anhand von bekannten Mustern Angriffe zu erkennen. Ergänzend dazu beschleunigt KI automatisierte Reaktionsprozesse auf Vorfälle, indem sie Warnungen aus verschiedenen Quellen kombiniert und automatisierte Maßnahmen wie die Blockierung verdächtiger IP-Adressen einleitet. Nach Angriffen unterstützt sie die Ursachenanalyse und hilft, Schwachstellen zu identifizieren.

Ein weiterer Schwerpunkt liegt auf der Phishing-Erkennung, bei der KI durch Natural Language Processing (NLP) verdächtige Muster in E-Mails analysiert. Solche Systeme identifizieren potenzielle Bedrohungen anhand von Dringlichkeitsausdrücken, ungewöhnlichen Absendern oder allgemeinen Anreden. Parallel dazu kommt Deep Learning zum Einsatz, um betrügerische Inhalte wie gefälschte Websites oder Deepfake-Medien zu erkennen. Diese Technologien analysieren spezifische Merkmale wie Domain-Namen, Layouts oder Audio- und Videodaten, um verdächtige Inhalte zuverlässig zu entlarven.

Auch im Bereich der E-Mail- und Social-Media-Monitoring-Systeme zeigt sich die Stärke der KI. Diese Systeme lernen kontinuierlich aus vergangenen Angriffen, um neue Bedrohungen frühzeitig zu identifizieren. Auf sozialen Plattformen können sie gefälschte Profile und verdächtige Interaktionen erkennen, was die Sicherheit sowohl für Einzelpersonen als auch für Organisationen erhöht.

Zusätzlich hilft KI, Insider-Bedrohungen zu minimieren, indem sie Verhaltensmuster von Mitarbeitenden analysiert und untypische Aktivitäten, wie unberechtigten Zugriff auf sensible Daten, erkennt. Gleichzeitig verbessern personalisierte Schulungsprogramme die Sensibilität gegenüber Social Engineering Bedrohungen. Mithilfe von KI werden Schwächen der Mitarbeitenden identifiziert, um gezielte Trainings und simulierte Angriffe bereitzustellen, die den Lernprozess gamifizieren und praxisnah gestalten. [13]

C. Persönliche Bewertung der Verwendung von KI im Bereich Social Engineering

Social Engineering bleibt eine zentrale Bedrohung, da es gezielt menschliche Schwachstellen ausnutzt, die selbst durch modernste technische Schutzmaßnahmen nicht vollständig eliminierbar sind. Besonders die Integration von KI in Social Engineering eröffnet neue Dimensionen für Angreifer, etwa durch die Erstellung realistischer Inhalte oder die Automatisierung von Angriffen. Gleichzeitig bietet KI jedoch auch erhebliche Potenziale für defensive Maßnahmen, wie die frühzeitige Erkennung und Abwehr von Angriffen. Durch Technologien wie Natural Language Processing oder Deep Learning können verdächtige Aktivitäten identifiziert und proaktive Gegenmaßnahmen ergriffen werden. Gleichzeitig ist es genauso wichtig, organisatorische und menschliche Aspekte gleichermaßen der technischen zu betrachten. Neben der kontinuierlichen Weiterentwicklung von Sicherheitstechnologien ist die Sensibilisierung und Schulung der betroffenen Personen wichtig, um das Bewusstsein für mögliche Angriffe zu stärken.

Abschließend soll deutlich werden, dass die Herausforderung in der Balance zwischen der Nutzung von KI für Sicherheitszwecke und ihrer potenziellen Gefährdung durch Missbrauch liegt. Nur durch Zusammenarbeit zwischen Technik, Psychologie und Weiterbildung kann die Bedrohung durch Social Engineering effektiv bekämpft werden, während gleichzeitig die Chancen der KI für eine sicherere digitale Zukunft genutzt werden.

REFERENCES

[1] "Social Engineering – der Mensch als Schwachstelle," Bundesamt für Sicherheit in der Informationstechnik. https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Cyber-Sicherheitslage/Methoden-der-Cyber-Kriminalitaet/Social-Engineering/social-engineering_node.html

[2] Z. Wang, L. Sun, and H. Zhu, "Defining Social Engineering in Cybersecurity," IEEE Access, vol. 8, pp. 85094–85115, May 2020, doi: https://doi.org/10.1109/access.2020.2992807.

[3] Harl. (1997). People Hacking: The Psychology of Social Engineering. [Online]. Available: http://www.textfiles.com/russian/cyberlib.narod.ru/lib/cin/se10.html

[4] K. Mitnick. My first RSA Conference. SecurityFocus. Apr. 2001. [Online]. Available: http://www.securityfocus.com/news/199

[5] Ponemon Institute LLC and Accenture. (Mar. 2019). Ninth Annual Cost of Cybercrime Study. [Online]. Available: https://iapp.org/media/pdf/resource_center/accenture_cost_of_cybercrime_study_2019.pdf

[6] Alahmed, Y et al. (2024). Exploring the Potential Implications of AI-generated Content in Social Engineering Attacks. 2024 International Conference on Multimedia Computing, Networking and Applications (MCNA). Available: https://doi.org/10.1109/MCNA63144.2024.10703950

[7] Fakhouri, H et al. (2024). AI-Driven Solutions for Social Engineering Attacks: Detection, Prevention, and Response. 2024 2nd International Conference on Cyber Resilience (ICCR). https://doi.org/10.1109/ICCR61006.2024.10533010

[8] 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

[9] Collier, H (2024). AI: The Future of Social Engineering! European Conference on Cyber Warfare and Security. https://doi.org/10.34190/eccws.23.1.2117

[10] H. Aldawood and G. Skinner, "An advanced taxonomy for social engineering attacks," International Journal of Computer Applications (0975 – 8887), vol. 177, no. 30, Jan. 2020, [Online]. Available: https://www.researchgate.net/profile/Hussain-Aldawood/publication/338623330_An_Advanced_Taxonomy_for_Social_Engineering_Attacks/links/5e20357b458515ba208aea83/An-Advanced-Taxonomy-for-Social-Engineering-Attacks.pdf

[11] Krombholz K, et al., Advanced social engineering attacks, Journal of Information Security and Applications (2014), http://dx.doi.org/10.1016/j.jisa.2014.09.005

[12] Schmitt, M., Flechais, I. Digital deception: generative artificial intelligence in social engineering and phishing. Artif Intell Rev 57, 324 (2024). Available: https://doi.org/10.1007/s10462-024-10973-2

[13] M. I. Khan, A. Arif, and A. R. A. Khan, "AI's revolutionary role in cyber defense and social engineering," jurnal.itscience.org, Oct. 2024, doi: 10.47709/ijmdsa.v3i4.4752.

[14] "'Red October' Diplomatic Cyber Attacks Investigation," securelist.com. https://securelist.com/red-october-diplomatic-cyber-attacks-investigation/36740/

[15] F. Mouton, M. M. Malan, L. Leenen and H. S. Venter, "Social engineering attack framework," 2014 Information Security for South Africa, Johannesburg, South Africa, 2014, pp. 1-9, doi: 10.1109/ISSA.2014.6950510.