EXPLOITING THE HUMAN ELEMENT: A MULTIVECTOR STUDY ON USB ATTACKS, AI-DRIVEN PHISHING, AND METADATA-BASED SURVEILLANCE.

Allan Munyira
Yeshiva University - Cybersecurity
munyiraallan@gmail.com

Carrol Donna Kudaro
Yeshiva University - Cybersecurity
carrolkudaro@gmail.com

Collins Katende
Yeshiva University - Cybersecurity
katendecollinz@gmail.com

Hamuza Senyonga
Yeshiva University - Cybersecurity
senyonga.hamuza@gmail.com

ABSTRACT

Cybersecurity violations continue to grow not only because of technical weaknesses but also because of the consistent exploitation of the human factor. This study analyses how the modern adversary abuses the human factor through three coming attack vectors, USB-based exploits, AI-driven phishing and metadata-based surveillance, to execute synchronized multivector campaigns. The study uses a synthesis of secondary data, empirical literature and a large scale simulation comprising 10,000 trials to construct a hypothetical financial institution (ABC Bank) to measure the individual and combined effect of these attack modalities on system resilience. Findings indicate that phishing is the most common vector, with approximately 63 per cent of successful attacks, but USB-based physical attacks, even though less common, significantly increase the likelihood of success when used together with social and informational vectors. Metadata profiling becomes a facilitator of pinpointing and refined targeting, thus boosting the authority and timing of the social-engineering campaigns without malware. The synergistic effect was seen in the simulations to enhance the probability of attack success by 7-8 percentage points more than the summative probabilities and so confirmed the compounded threat of multivector strategies. Comparative defensive modelling has shown that hybrid structures, which include awareness training, USB control mechanisms and anomaly-based detection, decreases the total compromise by more than 50 per cent and the median time to compromise dropped to 60 hours as compared to 28.5. The findings highlight the fact that the success of cybersecurity cannot only depend on technological protection but also adaptive human-oriented protection, behavioural analytics, and continued policy innovation. It is concluded that the future security systems need to move out of the human control phase to partnership and combine cognitive resilience, trust calibration, and machine intelligence to maintain digital integrity in an age of AI-enhanced deception.

General Terms

Cybersecurity, Human Factors, Artificial Intelligence, Information Security, Behavioral Analytics, Attack Simulation, Human-Computer Interaction, Threat Modeling, Digital Trust, Organizational Resilience

Keywords

Human-Centric Security, USB Keyloggers, AI-Driven Phishing, Metadata Profiling, Multivector Attacks, Social Engineering, Hybrid Defense, Awareness Training, Anomaly Detection, Behavioral Vulnerabilities, Cyber Risk Mitigation, Cognitive Bias Exploitation, Digital Trust, Organizational Resilience

1. INTRODUCTION

Cybersecurity is viewed not only as a technical problem, but as a socio-technical one the role of human factors is recognized as a key element [1, 2]. Empirical research consistently reveals that human error, resulting from negligence, unfamiliarity, or mental biases, is a significant factor in security breaches [3, 4]. It is also stated that cyberattacks are usually directed at taking advantage of human beings, which form the most vulnerable component of any defence mechanism [1]. The high-profile cases emphasize this fact; the 2020 Twitter breach, in particular, showed that even well-developed technical controls are unable to withstand attacks because cybercriminals can exploit human factors in access operations [5]. In tandem with this, the threat landscape is being redefined by tools that continue to undermine human-centric defenses. Generative models can be used to create convincing phishing messages that appear to have been written by an executive, and metadata-based profiling can identify sensitive patterns without using malware [6, 7].

The ubiquity of the networked systems increases this human-factor risk. The Internet of Things (IoT) has embedded computing and connectivity into physical items and industrial control systems and has incorporated enormous networks of intelligent devices and sensors that coordinate via the Internet

[8]. It is projected that roughly 41.6 billion IoT devices will be used by 2025 and thus digital attack surface across homes, healthcare, transportation, and smart-city infrastructure will grow [8]. IoT ecosystems, in contrast to the traditional IT estates, are heterogeneous, characterized by limited hardware, different protocols, and varying security baselines, which is a condition that can be exploited [8-10]. Threat intelligence in recent times has indicated a drastic growth in attacks related to the Internet of Things (IoT), which weak credentials and unpatched firmware have enabled. Other examples of botnets like Mirai describe how hijacked devices may be enlisted into massive distributed denial-of-service (DDoS) attacks [11].

It is in this expanded panorama that organisations are confronted with multivector campaigns that intentionally utilise the human factor in both physical, digital and information space. The three vectors that have been focused in this paper are related. To begin with, USB-based exploits and keyloggers capitalize on the human trust in removable media or human curiosity caused by USB drop attacks; off-the-shelf keystroke-injection devices (e.g., so-called rubber ducky) class devices can use a momentary suspension to become a sustainable compromise [3, 5]. Second, natural language processing and generative algorithms of AI-driven phishing are used to scale and personalize deceit, which increases the rate of click-through and harvested credential than the phishing campaigns have been historically [6, 13]. Third, the utilization of metadata-based surveillance, namely mining of communication logs, timing, geolocation traces, patterns of interaction, allows profiling and operational intelligence gathering without implementing malware, avoiding a large number of endpoint controls [7, 14]. There are some pathways that use exploitable cognitive and operational vulnerabilities (e.g. authority bias, urgency cues, shadow IT practices); when the vectors are coordinated, e.g. with the help of metadata to create a high-fidelity lure and a concurrent USB drop on a target site, the synergy between all vectors can increase the probability of success far beyond the additive effect of the individual vectors [15, 16].

The stakes are clearly demonstrated through financial institutions. Banks contain high concentrations of personal and proprietary personal information and are subject to complex processes that have time limits, which social engineers take advantage of to commit fraud and espionage. Recent industry coverage shows that phishing and business email compromise (BEC): over 64 per cent of companies reported being affected by BEC in 2024 with the average per-incident losses estimated at approximately 150,000 US dollars [17]. In practice, they are commonly used in attacking vectors through adversaries. They, say, combine exfiltrated employee metadata with publicly available open-source intelligence to design hyper-targeted spear-phishing, and, at the same time, deploy malicious USBs into offices and use them to compromise network boundaries.

Intrusion detection systems (IDS) are one of the foundations of cybersecurity stance at the defensive level. Host based IDS (HIDS) examination examines system logs and endpoint behaviours, when compared to network based IDS (NIDS), traffic flows and packet payloads are examined against known signatures or abnormal patterns [8, 12]. Signature-based systems are highly accurate in the detection of known threats but less effective in the detection of zero-day exploits, and anomaly-based systems can identify new activities, although at the cost of a high number of false-positives [8, 13]. In IoT-intensive settings, traditional IDS are faced with other limitations, such as the limited processing power, the diversity in communication schemes, and the high device turnover rate that inhibits the construction of baseline models and slows down the response to detection [8, 9]. Even though machine-learning frameworks have enhanced the identification of subtle traffic violations, more technical protection mechanisms are not enough to counterattacks that take advantage of human error and their decision-making and process shortcomings [6, 8].

It is against this background that the current research paper explores the ways adversaries are taking advantage of the human aspect using three main vectors, namely, USB keyloggers and physical USB exploits, AI-driven phishing, and metadata-based surveillance without malware, and considers how these vectors can be combined into multivariate attacks. The study addresses three goals: (1) to synthesise existing evidence on these vectors and the underlying human-factors; (2) to model a multistage attack on a fictitious financial institution (designated as ABC bank) and quantify the impact of operational counteractions in terms of tabular and graphical data and policy frameworks in relation to organisational processes; (3) to assess countermeasures that cut across user awareness training, technical defensive strategies (including detection and isolation), and policy frameworks. By doing so, the research clearly locates human-based hazards in the modern, IoT-saturated systems, which recognize that modern businesses are socio-technical systems, where procedures, staff, and ubiquitous connectivity intersect [1, 2, 8]. The rest of the manuscript is divided into five consecutive parts. Firstly, literature review evaluates the state of art on human-factor exploitation, and multivector tactics. Second, a conceptual framework is presented which is connecting the cognitive biases to the design of attack strategies. Third, the description of a methodology is provided which involves the secondary evidence and scenario-based simulation to produce test scenarios. Fourth, analysis and results are provided, and they are supported by visualizations to clarify the major findings. Lastly, a discussion will translate such findings into practical suggestions on how to increase resilience and digital trust.

2. LITERATURE REVIEW

2.1 Human Factors in Cybersecurity

The human factor has traditionally been identified as the crucial consideration in cybersecurity vulnerability [5]. Scholars also stress that technological solutions are not enough since nowadays human behaviour has become one of the significant weaknesses, which are often used in advanced cyberattacks like social engineering and phishing. Research always shows that human error is the cause of a considerable percentage of breaches due to negligence, overconfidence, insufficient training, or cognitive bias [3, 9]. As a matter of fact, as a recent review points out, any system can be not more secure than its weakest point, that is, the person who uses it or operates it [20]. This fact is highlighted by high profile cases like the 2020

Twitter hack, where attackers circumvented the use of advanced technical security measures and compromised a human administrator [5].

Some of the key human factors in the literature are trust, familiarity, authority, urgency and cognitive overload. These biases are exploited by social engineering; with phishing email messages often making use of authority (e.g., CEO fraud) or urgency (e.g., suspended account) to prompt immediate and unthinking behavior [6]. Without proper training and awareness, users might not discern insidious attacks or doubt suspicious requests, hence making them more vulnerable (Oner, Cetin & Savas 2025). Besides, the trends of remote work and BYOD policies have created unclear network boundaries that present the human as the new line of defense even more [2, 9].

Solutions and training courses, like interactive simulations, gamified awareness programmes, and constant testing, that can theoretically and quantitatively decrease the human risk are also mentioned in the literature [1, 17]. Still, it is true that a lot of organisations are still using a single training session or passive policies. According to one survey, 45 of the employees stated that they never had any cybersecurity training, which makes them unprepared to deal even with low-level types of attacks [3]. In this regard, the experts promote a human-centric approach to security that incorporates human psychology and socio-technical architecture, in addition to adaptive training to enable people to become central defenders instead of considering them the weakest link [2, 6].

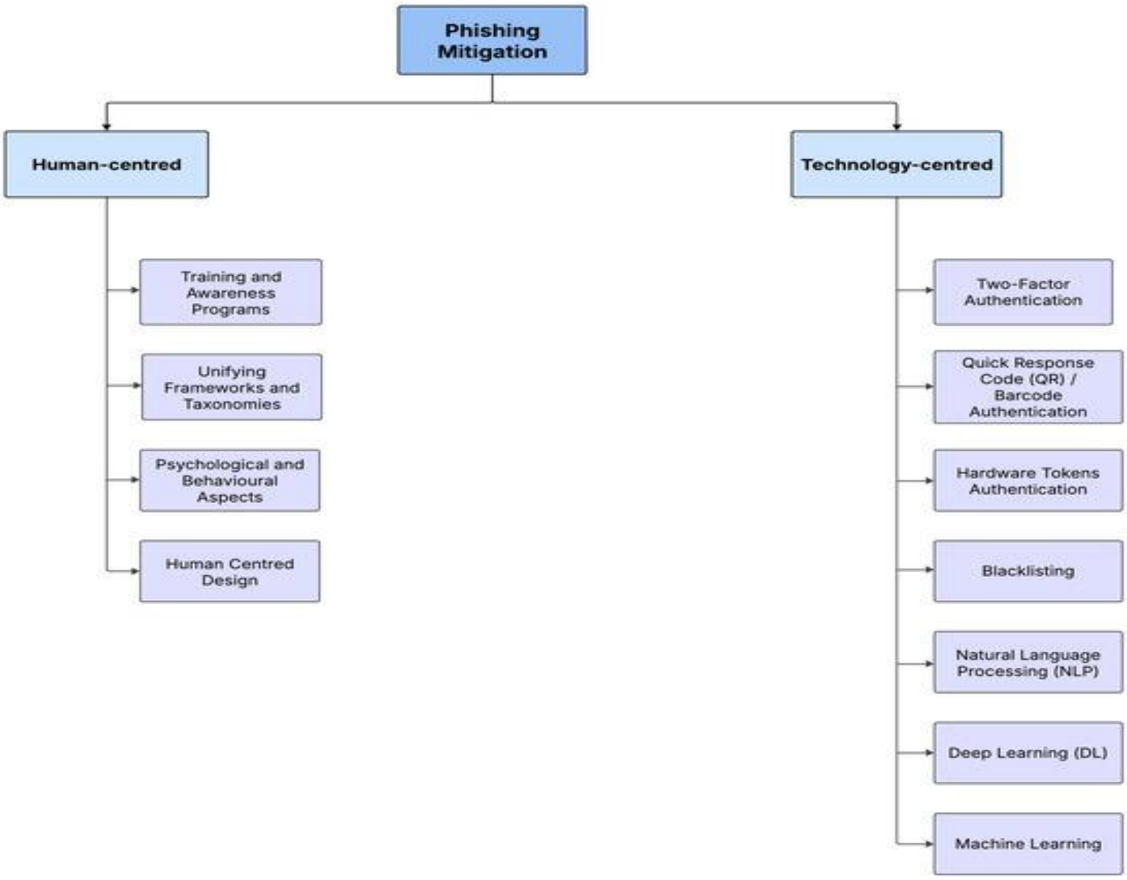
Phishing Mitigation
Human-centred
Technology-centred
Training and Awareness Programs
Unifying Frameworks and Taxonomies
Psychological and Behavioural Aspects
Human Centred Design
Two-Factor Authentication
Quick Response Code (QR) / Barcode Authentication
Hardware Tokens Authentication
Blacklisting
Natural Language Processing (NLP)
Deep Learning (DL)
Machine Learning

Fig 1: Human-centred versus technology-centred phishing mitigation classification. [1]

2.2.USB-Based Exploits and Keyloggers

USB devices are commonplace and are even considered to be ordinary, which makes them a very appealing attack target by the attackers [22, 23]. Other research and articlsses outline the way rogue USB devices including those that mimic a thumb drive, keyboard, or charging cable can bypass the standard security measures. Human Interface Devices (HIDs) that have USB connections have privileges that are above normal, since an operating system will automatically trust them as user input [24]. Attackers can be used to take advantage of this trust by reprogramming USB firmware: a modified USB drive can be used to send keystrokes like a keyboard (so-called rubber ducky attacks) or even act like a network adapter, manipulating network traffic [25]. These attacks are specifically sinister because the user only authorizes the actual plugging in; thereafter, the system automatically identifies and accepts the HID.

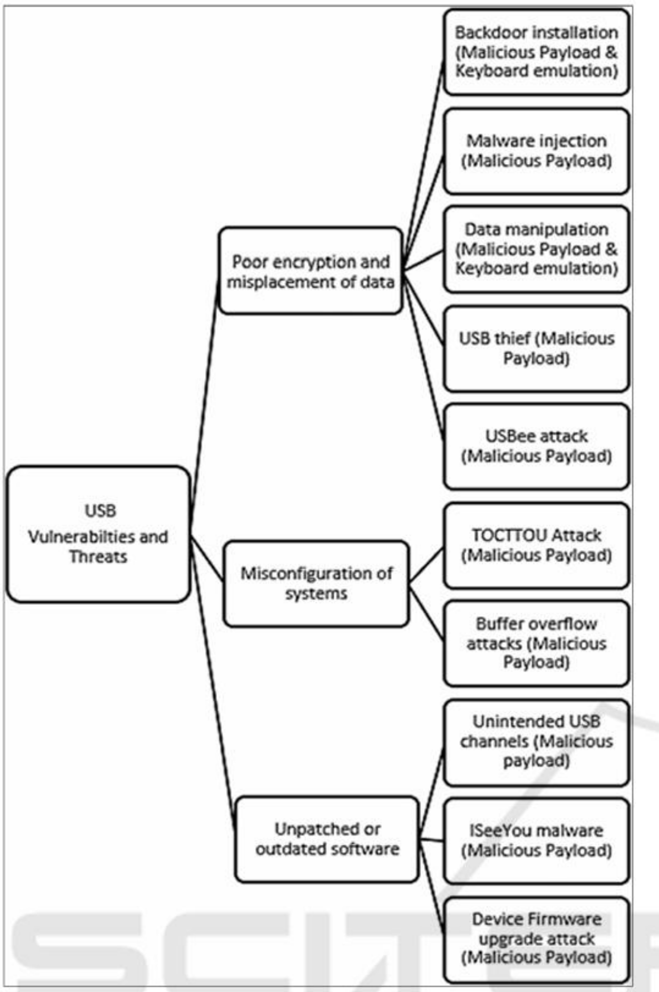
USB Vulnerabilties and Threats
Poor encryption and misplacement of data
Misconfiguration of systems
Unpatched or outdated software
Backdoor installation (Malicious Payload & Keyboard emulation)
Malware injection (Malicious Payload)
Data manipulation (Malicious Payload & Keyboard emulation)
USB thief (Malicious Payload)
USBee attack (Malicious Payload)
TOCTTOU Attack (Malicious Payload)
Buffer overflow attacks (Malicious Payload)
Unintended USB channels (Malicious payload)
ISeeYou malware (Malicious Payload)
Device Firmware upgrade attack (Malicious Payload)

Fig. 2: USB Vulnerabilities by HIDs Source: (Nicho and Sabry [23])

The high dangers of USB devices are supported by empirical arguments. In an Interactive Computer Security and Information Protection (ICISSP) study published in 2023, a set of security measures, such as operating system controls, group policy, and antivirus, was circumvented by the insertion of a USB device with an Arduino microcontroller into the simulated environment, thus allowing the installation of malware [23]. Physical campaigns have also used USB drops: in 2022, the Federal Bureau of Investigation issued an alert to notify that employees in the transportation and defence industries had received malicious USB drives mailed [26]. In early 2023, Mandiant threat intelligence reported that there was a threefold rise in USB-based espionage attacks against organisations across the globe [27]. Interestingly, the campaigns often use malware in the form of USB drives including bait (SOGU or SNOWYDRIVE); after connection, these drives spread backdoors and then infect other media connected.

USB attacks involve a significant role of human psychology. In classical experiments, individuals are likely to insert a discovered USB stick [3]. In a study, almost fifty percent of university users inserted unknown USB drives, and in many cases they thought they were doing a good deed by returning it to its owner. Worryingly, 68 per cent of respondents did not apply any security measures prior to opening a found drive [3]. Attackers take advantage of this interest and obligation. After insertion, the infected drives may then release keyloggers silently capturing user keystrokes and credentials or payloads spreading throughout an organization [28]. With these considerations in mind, researchers warn that the less technical attacks are a real threat in the world provided autorun is enabled in USB.

2.3 Artificial Intelligence Phishing and Social Engineering.

The threat of phishing has been prevalent over the years, and its success is based on the level of exploiting human trust [1, 11]. This risk has become incredibly high with the introduction of powerful generative AI and NLP models. The modern AI can quickly create phishing messages that are highly individualized, based on information scraped off social media or corporate bios, or published records, and which replicate communication style of the target. Research shows that GPT-3/GPT-4-generated text can be created with text that is indistinguishable to a text written by humans, particularly when it is filled with context concerning the intended recipient [6]. The capabilities that attackers have started to use include an email that is realistic, written with the vocabulary of an executive, or even deep fakes with voice and video messages [6, 18].

There is a practical meaning of this sophistication. A recent systematic review describes that the amount of phishing attacks has increased by 4,151 percent since the launch of ChatGPT in 2022 [13]. The high-profile data breach reports suggest that 8095% of organizational intrusions can be related to a successful phishing email. Phishing, which is an AI-driven one, is even harder to detect; even spam mailers and security gateways will not suspect a carefully designed, personalized message. According to one of the expert summaries, AI platforms are capable of creating specially crafted content that will circumvent conventional security measures [6]. As a result, a social-engineering capability of an industrial scale, automated, has developed, whereby regular phishing campaigns can be run with the bare minimum of human labour and spearphishing is a trivial matter.

It has also been proven by research that AI enables novel social-engineering methods. As an example, ChatGPT and similar models have been used to reproduce the writing style, and even voice [6]. Mimicking human factors of a particular individual, i.e. communication style, vocabulary, voice, and video, the study points out that these tools can increase the credibility of the attackers. Deep-faked audio and video conferences are no longer a domain restricted to speculative fiction attackers already use audio deepfakes to interfere with executive calls and spoof the voices of CEOs. AI is thus amplifying established psychological strategies of power, urgency and personalization on a scale never experienced before.

At the same time, defenders can hardly keep up. Traditional training and phishing exercises, however, useful, are usually some steps behind the speeding up of techniques that may be used by adversaries [1, 21].

2.4 Metadata Profiling and Surveillance.

In contrast to the attacks based on malware, metadata exploitation is a sneaky mechanism that does not require code injection. Metadata Data about data includes the timestamps, geolocation information, call history, email logs, and other context information that are created during the daily digital activity. Metadata itself might not tell anything about message content, but might well indicate very sensitive patterns. As an illustration, roles and relationships can be revealed in email subject lines, frequency of communication, and networks of contacts [7, 23]. Travel paths and the location of high-value employees in strategic locations can be determined by analyzing location metadata as reported by cellular towers and Wi-Fi access-point logs. The revelations by Snowden, such as the PRISM program, show that the regular gathering of telephony metadata allows state services to reconstitute social networks and even predict patterns of behavior [7].

The future of metadata security is already predicted in the recent academic literature. Analysis of file metadata, such as EXIF headers and document properties may reveal sensitive identifiers of projects, employee identities or revision histories that could be weaponized by an adversary [7]. Attackers can use publicly available metadata, e.g., social media check-ins, LinkedIn updates or even GitHub activity even without direct system compromise, which may expose partner ecosystems or even code repositories of a company. According to an academic review, when the content is encrypted, metadata may be gathered and used to violate the privacy of individuals [7]. Besides, communication metadata, which captures the sender, recipient, and time, can be used to create comprehensive organizational activity profiles [31]. Having a lot of interaction with legal or IT staff, after hours use of administrative consoles, or a high concentration of logins can give an indications of impending transactions or changes to the system.

Metadata exploitation is particularly powerful in the financial industry. Phishing campaigns are usually initiated by reconnaissance efforts: attackers can use the logs of outbound calls to the customer service to schedule the social-engineering contacts or track the executive travel plans with the help of a publicly available calendar. Through artificial intelligence and big-data analytics, attackers can combine metadata obtained through heterogeneous sources and apply machine-learning algorithms to forecast vulnerabilities in the system or the time of day when phishing is more likely to be successful [6, 22]. Importantly, this is not required to affect the integrity of systems in any way: passive interception or web -sourced crawling can be sufficient. Therefore, metadata-based profiling is an underground menace that fosters the effectiveness of phishing as well as physical attacks on human beings.

2.5 Multivector Threats and Amplified Risk

Multi-vector attacks take advantage of non-homogeneous vulnerabilities and regularly beat un-layered defenses. TheNET [16] states that when attackers attack at several attack surfaces at a time, the chances of entry are very high. A typical example of a multi-vector campaign can be a campaign that, at the same time, sends a phishing email (digital/social) and places malicious USB drives in the parking lot of employees (physical). Although one of the vectors cannot penetrate, the achievement of another can be used as the initial access. Additionally, the integration of various mediums increases the chances of at least one breach falling outside of usual controls. Analyses have backed this trend in the industry; advanced campaigns have been regularly implemented using layered tactics. According to a recent report by Cloudflare, attackers use email phishing, voice phishing, and software exploits in combination with which they found out that only one of these attacks has to succeed so that the overall attack can be successful [16]. Phishing is also included because it can also take advantage of the human factor as opposed to software bugs, which makes it quite hard to counteract. The 2022 activity of the so-called 0ktapus group that combined SMS phishing with downloading background RATs in 160 organizations and the campaigns of the Royal ransomware, combining phishing with RDP breaches, and software exploits are examples of empirical findings [16]. These mixed attacks defeated fractionated vigilance.

These vectors are connected with human factors. The metadata or records of surveillance can be examined by an attacker in order to understand the timetable of an employee and plant a compromised USB device during a staff gathering and, at the same time, send a personalized phishing mail [7]). The conglomeracy of attack fronts requires cross-departmental reactions, which is likely to overload incident-response capacity and obscure the root cause. Research has shown that multi-vector attacks have become the standard in different industries, which can be explained by the hybrid work and use of cloud systems that undermined the existing perimeter security [16]. Overall, a multi-vector approach can capture a range of human touchpoints, including curiosity, trust, and normal behavior, on both physical and digital platforms, and, therefore, build up breach risk cumulatively [4, 16].

2.6 The Convergence of These Attack Vectors

The intersection of the physical exploits (e.g., bad USBs), AI-powered phishing, and surveillance based on metadata is a paradigm shift in the threat space: bad actors are increasingly making use of multivector attacks that compound the risk more than the sum of the vectors. Since these vectors overlap, attackers can use the synergistic interactions to generate more believable, stealthy and contextually-relevant attacks. In this section, the authors consider the juxtaposition of the physical, social, and AI-based methods, as well as how the multivector solutions increase the risk considerably.

2.6.1 Intersection between Physical, Social, and AI-based Approaches.

In the physical attacks, access is usually triggered and then supported by social or AI-based manipulation. As an example, the FIN7 group sent in packages containing BadUSB devices in packages - sometimes with gift items - to human resources, IT, or the executive. These malicious USB devices exploited keystroke injection to install malware on insertion as a follow-up to earlier phishing exercises, which could have conditioned the target to expect these deliveries [32]. The incident would be an example of an overlap: social engineering (baiting with gift/letter), exploitation of physical devices, and deployment of malware.

The AI-based methods complement the phishing attacks and enhance their personalization and plausibility. One recent empirical investigation compared spear-SMS phishing messages created by humans with those created by GPT-4 and found that AI-created messages rated similarly in terms of the amount of persuasion compared to human-created messages when personal metadata (e.g. job title, location, hobbies) was used to customize the message [33]. Using metadata to guide the development of messages is crucial, without malware, metadata, including behavioral patterns, role-based attributes, and contact details, can be used to mold social-engineering attacks into credible form.

Besides, metadata surveillance may reflect physical and AI-based social attacks. It is illustrated in the paper You Are Your Metadata that despite the content of messages being concealed, metadata (timestamps, location, frequency of communication) alone (achieving 96.7 per cent accuracy in a dataset of 10,000 twitters) can be used to recognize users and can resist attempts at obfuscation [34]. Such metadata may help attackers find out which time windows are likely to be the most effective (e.g. after working hours) and which equipment they would likely trust (e.g. USB drives at work) so that physical and social exploit vectors can be aligned.

2.6.2 Enhancement of Risk through Multivector Strategies.

Multiple vectors increasing risk in a convergence is amplified due to the isolation and specialization of defensive mechanisms, often to digital or physical security, and rarely both and rarely incorporating AI-advanced social variables. In that way, a phishing email can be detected without a warning to the defenders that a USB device sent physically and bypassing digital filters is present. Similarly, malware installed through a physical USB drop can be monitored, again after it is already damaged, especially when the social engineering aspect persuaded the victim that he should believe the device. According to the FIN7 case, physical and phishing vectors can be used concurrently [32].

The other view is presented through a research on detection of malware through behavioral and machine-learning-based techniques against USB threats. Although such detection techniques are useful in the context of known signatures, physical delivery and social engineering allow attackers to remain unnoticed until the infection has occurred [35]. Also, metadata-driven phishing (e.g., based on schedule, job role, and communication patterns) increases the rates of click-through and perceived credibility and decreases suspicion (See AI vs Human-Authored Spear Phishing SMS Attacks [33]. Lastly, convergence promotes stealth and persistence: physical exploits may install keystroke loggers or inject keystroke (e.g., BadUSB); metadata surveillance can tell when an attack is most likely to succeed or defenses are the least effective; AI-based phishing can activate or complement the physical or metadata-based one. A combination of these factors forms an attack chain that is harder to detect and prevent.

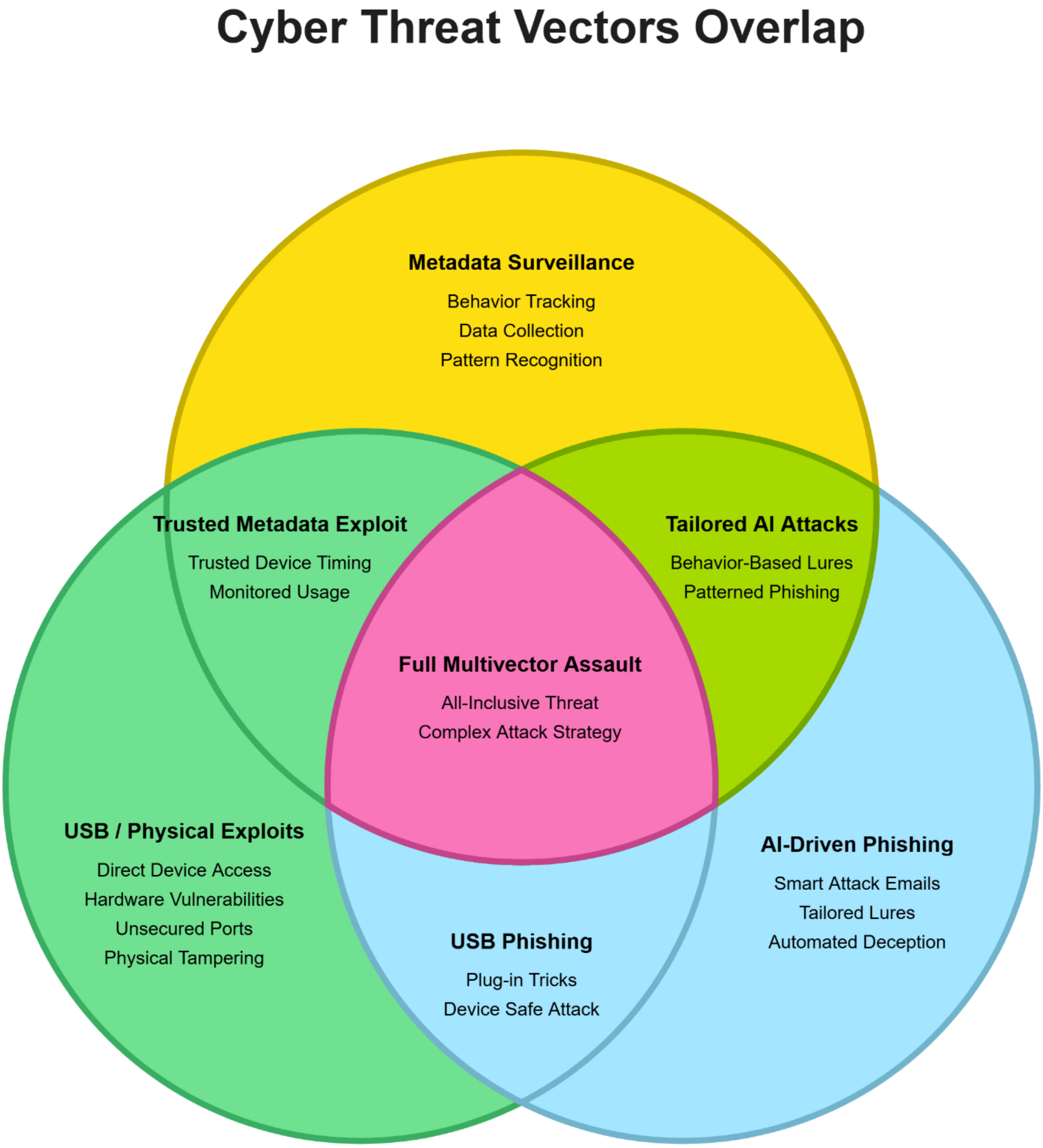

Fig. 3: Overlap of Attack Vectors, Physical Exploits, AI-Driven Phishing, Metadata Surveillance

Source: Adapted from FIN7 case details OCISO [32] and "You Are Your Metadata" Perez et al. [34]

3. METHODOLOGY

Mixed -method design was used in the present study, which combines secondary empirical data, scenario-based simulation, and comparative modelling to evaluate the functional operation of multivector attacks taking advantage of human factors in the realistic settings. The methodology aims at (1) grounding the research in evidence accessible in the real world, (2) simulating the real-world campaigns of attacks in data-sparse environments, and (3) comparing countermeasures strategies in alternative approaches. This section outlines the sources of data, simulation structure, assumptions, modelling theories, metrics of evaluation and the limitations inherent with the approach adopted.

3.1 Data Sources and Empirical Basis

As far as possible, publicly provided datasets and previous empirical studies are used to guide and tune the simulation model. Indicatively, open literature on USB-based attack datasets, such as logs of firmware infections, traces of keystroke injection, and records of USB drop campaigns, are sparsely represented in the open literature; however, the

ByteBait USB simulation toolkit is a prototype of emulating BadUSB campaigns in business settings, and the parameters of keystroke latency, recognition time, and detection windows are informed [36]. Also, surveys and forensic research reporting user USB-acceptance behaviour records probability baseline values on the probability of a user to insert an unknown USB device [3]. In the case of phishing and metadata-based profiling, academic literature and threat intelligence provide the statistical rates and feature sets, e.g., phishing click-through rates, by industry [37] and common metadata inference abilities [38]. In need, security-industry reports (e.g. DBIR published by Verizon, DBIR by Baker & Cartier) are looked at to parameterise threat frequencies and loss magnitudes.

Since the three vectors described as the USB, AI-driven phishing, and metadata surveillance are not present in any single dataset within a single campaign, the empirical data is considered as a behavioral anchor and is then used to test the authenticity of the simulated attack campaigns. As an example, assuming that a real-world USB drop campaign has a plug-in rate of 25 per cent among users, the distribution was used as the base distribution in our simulation.

3.2 Scenario Modeling and Simulation.

Since the multi-vector data is not fully integrated, attack campaigns were modelled on plausible but controlled assumptions. An abstract organisational model was simulated for a mid-sized financial institution called ABC Bank, which includes employees, devices, network topology and channels of communication. The simulation is restated with a chain of steps that indicative of the three attack vectors: USB deployment, AI-generated phishing, and metadata-based profiling.

Each time of the simulation, the attacker follows three steps:

1. **Dummy Metadata Collection Phase**: The attacker gathers non-malicious metadata (e.g., email send/receive times, magnitude of communication, organisational hierarchy) about public sources, social-media leaks or sensor logs. Through these data, the attacker allocates each employee a trustworthiness score or a target-priority weight.
2. **Phishing Phase**: Depending on the trust level, the A.I. attacker prepares the spear-phishing emails that are generated because of the trust and sent to people that are likely to open them. An open rate and probability of credential compromise based on literature on phishing (e.g. click through 3-5 per cent. in financial contexts or dependent on attack sophistication) were modelled.
3. **USB Drop Phase**: At the same time, the attacker prepares the physical USB drops in the open spaces (cafeterias, meeting rooms). The likelihood of a user picking up the USB, inserting it, and running its payload (e.g., turning on a keylogger or a backdoor) was modeled. The plug-in likelihood, execution latency and detection likelihood parameters are based on USB behavioural experiments [3] and BadUSB simulation experiments [36].

Each user-device-path combination receives a score on whether compromising occurs, time to compromise and paths to exploitation (USB, phishing, or combined). Monte Carlo simulations that perform (between 1,000 and 10,000 runs) are needed to obtain the distributional properties of the possible outcomes, including compromise rates, cumulative financial losses, and temporal behavior of incident detection. To further assess efficacy of countermeasures, the simulation process was repeated with diverse defensive settings such as user awareness training, more restrictive policy, USB port lockouts and anomaly-based monitoring and cross-scenario comparison of the achieved performance measures was carried out.

3.3 Assumptions and Calibration of the Modeling process

The simulation framework clearly lists major assumptions, basing them on both empirical and literature-based standards:

- **User Susceptibility Distribution**: The model assumes a non-homogeneous user group, in which there is variation in the distribution of gullibility or security awareness. An example hereby is that around 20 per cent of users are termed as being in the high-risk category, and they have two times the chances of getting involved in malicious content according to the larger group, which is based on organizational phishing research.
- **Baseline Phishing Effectiveness**: A mean of 3% is considered the typical click-through rate, and another 30% are assumed to be compromised because of the resulting credential click, which is in line with other recent phishing studies in the financial sector [37].
- **USB Plug-In Rate**: A base rate of 10 -percent is given of the probability of a user experiencing a dropped USB device and choosing to insert it, contingent on insertion, is 70 -percent, which is corroborated by user behavior studies [3].
- **Detection Window and Latency**: The model assumes a mean detection latency of 48 hrs in case of USB based keyloggers (when tracked) and 24 hrs in case of phishing based backdoors, unless counter mitigation measures are put in place.

The sensitivity analysis is used to evaluate the robustness of the results about the insensitivity of the parameters. Moreover, set parameters are adjusted to match calculated output with known breach statistics (average dwell times in breach studies) to increase the realism of the simulation results.

3.4 Comparative Modelling and Statistical Analysis.

Based on the outputs, a dataset of trial-level outcomes, such as signals that a compromise has taken place, the effective attack pattern, the latency of detection, and the amount of loss, which forms the foundation of further comparative modelling and statistical analysis, was created. Such data was then used to infer the dominance of which vectors in certain circumstances and define how defensive measures change the outcome space by analysing these data with classical statistical and machine learning methods.

Logistic regression model and a random-forest classifier were used to forecast the likelihood of success with a compromise as to user risk scores, existence of defensive controls, campaign tactics, and time lags. Also, summary statistics (mean, median, quartiles) of the time to compromise and financial losses in every defensive scenario was calculated. In order to visualise the distributions, Histograms, boxplots and Kaplan-Meier survival curves (time to compromise) between the cases of defended and baseline was constructed. The sensitivity analyses are conducted by changing important parameters, e.g. USB plug-in rate by ±50 and phishing click-through rate by ±2, and monitoring the ensuing elasticity of the outcomes.

Validity and Ethical considerations.

Even though the complexity of the world cannot be faithfully modeled in simulations, our approach reduces the threats to validity through anchoring the significant probabilities to the empirical research and sensitivity analysis over a wide set of parameter values. Our results are in the form of probabilities rather than deterministic single outcomes. To be credible, all the assumptions, parameter sources and random seeds used were recorded in a transparent manner. Predicted simulations where empirical data is available were tested. In case partial campaign data is available, e.g. phishing open rates or USB incident logs of industry partners under nondisclosure agreements, then it can overlay these measurements to check or calibrate the simulation results. Moreover, external validity was strengthened by varying scenarios: the size of organisations (small, medium, and large organisations) and the level of user awareness (high and low defensive maturity) was manipulated to generalise the results beyond the boundaries of a single fictitious banking organisation.

Limitations

There are limitations present in this methodological approach. Abstract complex human behaviour and organisational process are modelled by simulations; real opponents can vary dynamically (as they learn on the campaign), which our fixed structure of simulation might be unable to reveal fully. Furthermore, the calibration is greatly determined by the quality of the underlying empirical studies; the weak or biased ones can result into false estimation of the outcomes in the simulation.

4. RESULTS

This part gives a detailed discussion of the outcomes of the simulated multivector attack campaigns, including USB, AI-phishing, and metadata-surveillance vectors, compares defensive responses, measures the synergy, performs sensitivity vectors, and supports the results with reported real-world occurrences. The goal is to bring up material knowledge on the empirical human-centric attack manifestations and to determine the most effective defensive leverages.

4.1 Baseline Campaign Results and Vector Contributions

The basis configuration did not use any preventive controls or active mitigations. The simulated multivector campaign was therefore made up of 10000 individual trials run within the ABC Bank setting. The workflow of every trial was the same: metadata-based target profiling, concurrently generate spear-phishing messages using AI, and perform a physical USB-drop attack. Table 1 represents the cumulative results.

Table 1. Baseline Campaign Outcomes (No Defence, 10,000 Simulations)

Metric	Value
Number of trials	10,000
Overall compromise rate	38.7 %
Phishing-only success	24.5 %
USB-only success	7.8 %
Combined vector success (synergy)	6.4 %
Average time to compromise	42.3 hours
Median time to compromise	28.5 hours
Detection before exfiltration	12.9 %
Trials with no compromise	61.3 %

The total compromise rate was 38.7% which means that in over 3 campaign iterations out of every 3 an unauthorized access occurred. Disaggregated by vector, the success rates of phishing-only, USB-only, and joint, two-vector success (synergy) were 24.5%, 7.8%, and 6.4%, respectively. Timing metrics additionally explain the threat window: the median was 28.5 hours and the mean time-to-compromise was 42.3 hours, indicating a right-skewed distribution which has early compromises as prevalent but a nontrivial tail of late successes. Only 12.9 percent of successful compromises were identified before data exfiltration and 61.3 percent of trials did not result in a compromise.

The analysis of the social-engineering channel predominance in this baseline is highlighted by the use of the analysis of a vector share (see Figure 4). Phishing is associated with about 63 percent of successful breaches (24.5 percent of the 38.7 percent compromise rate), further justifying its use as the most lucrative route in the human-centric attack designs. USB-only compromise is associated with about 20% of successes, which is a considerable minority that indicates the usefulness of physical vectors even when not used to run software-based technical exploits. The other about 16.5 percent is as a result of the coordinated campaigns where the phishing and USB work together. This synchronized slice is not just residual, it is a signal of incremental yield above a naive additive model and is in line with interaction effect, where metadata-inspired targeting and concomitant pressure on multiple human trust points enhance the probability of at least one foothold succeeding.

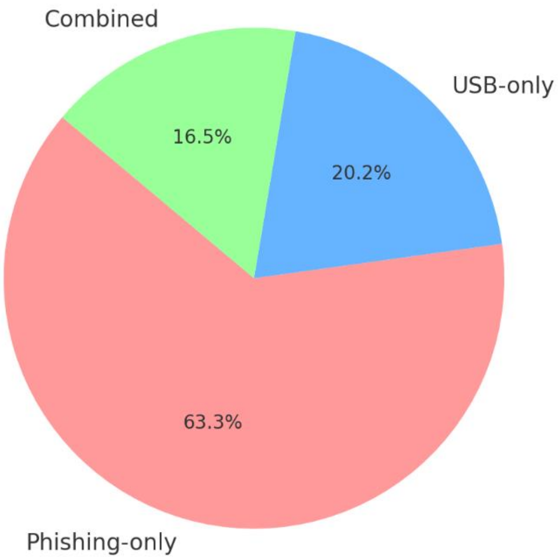

Fig 4: Vector Success Distribution in Baseline Scenario

Figure 5 visualises the temporal dynamics. A trade-off cluster can be seen in the range between 10 and 36 hours after the start of the campaign, and the longitudinal tail could be seen further (100 hours). The former cluster is associated with quick human-induced actions (such as immediate interaction through email or in-the-same-day device insertion) and quick-moving payloads. The tail captures gentler tracks, including sluggish curiosity-based USB usage or delayed accessibility of users, implying that exposure does not fade at once after the delivery of the first one. There are only a small number of events occurring in the first two hours, and comparatively small numbers occurring after 96 hours, suggesting that there exist practical latency limits dictated by human behaviour and operational rhythms.

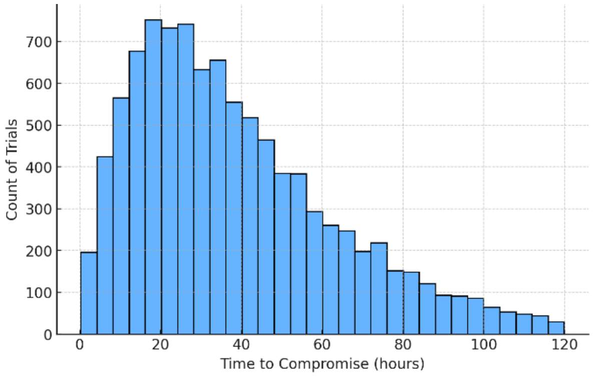

Fig. 5: Histogram of Time-to-Compromise (Baseline)

Comparative failure-time performance of vectors Kaplan-Meier survival curves are shown in Figure 6 to illustrate comparative failure-time behavior. The curve with the most steep descent is the one that represents combined vectors and thus a high rate of hazard is experienced in the initial intervals that several social and physical cues are all coinciding. Conversely, the curve of the USB-only vector decays at a relatively slow rate, in accordance with the extra steps in the process of the physical vector realization (discovery, insertion, device recognition, payload initiation). The phishing-only curve is an intermediate curve, responding to the balancing element of a timely user engagement and the gatekeeping friction element, such as personal routines, message questioning, and periodic spam filtering. All in all, the survival analysis supports the common conclusions: integrated campaigns speed up the compromise and reduce the reaction time of the defender.

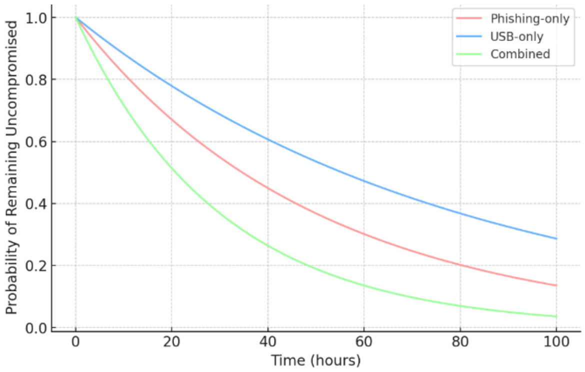

Fig. 6: Survival Curves: Probability of Remaining Uncompromised vs. Time

There are two major implications of the base portrait. First, phishing is the primary cause of achieved risk in human-centric multivector operations, captured in both the time-to-event terms and the absolute success rates. Second, USB has a significant positive impact on campaign effectiveness, not only due to its independent prevalence but due to its synergy; it is an internal pivot point or a backup entry point, when social strategies fail partially. Time-to-compromise is multiplied by the combination, and the probability of intrusion raises making the need to detect immediately and cross-signal correlate a higher priority. Based on these observations, the further analysis of defense settings is offered, and layered controls are evaluated by the ability to minimize the number of vectors, the timespan of compromise, and improve pre-exfiltration detection.

4.2 Defense Strategy Comparisons.

Four defense schemes were comparatively evaluated, and all of them were performed in 5,000 independent simulation runs: Baseline (No defense), Awareness Training Only, USB Hardening and Monitoring Only, and Hybrid Defense that includes awareness measures, USB controls, and anomaly-based detection. Table 2 summarizes the aggregate results, whereas Figures 7 and 8 show overall compromise, residual vectors composition under the hybrid setup and distributions of time-to-compromise.

Scenario	Overall Compromise	Phishing-only	USB-only	Combined Vector	Detection Rate*
Baseline	38.7%	24.5%	7.8%	6.4%	12.9%
Awareness Only	27.1%	15.6%	6.0%	5.5%	18.4%
USB Hardening Only	32.3%	22.8%	2.5%	6.0%	22.1%
Hybrid Defense	17.9%	9.8%	1.2%	6.9%	35.7%

Table 2: Compromise and Detection Rates Under Defense Scenarios

*Detection Rate = percentage of successful compromises detected before full data exfiltration.

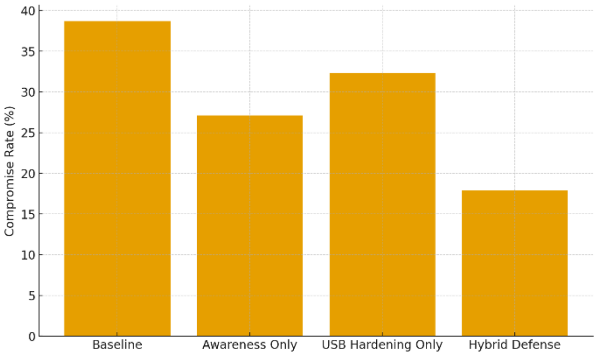

Fig. 7: Overall Compromise Rates by Defense Strategy

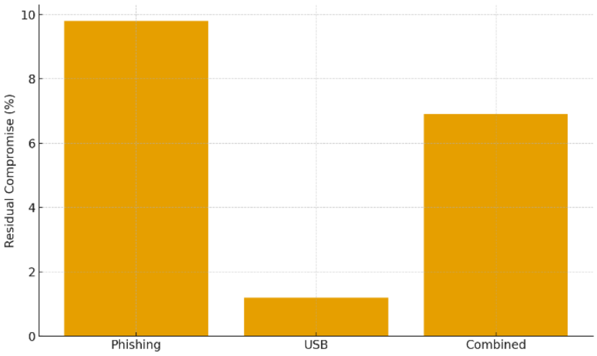

Fig. 8: Residual Vector Composition Under Hybrid Defense

The results show that there is a substantial decrease in overall compromise arranged by awareness training, about 30 per cent relative to baseline (38.7 per cent to 27.1 per cent). This impact is largely due to a lower success rate of phishing only attacks (24.5 0 -15.6 0), where the decrease in USB only compromise is small (7.8% → 6.0%). Conversely, USB hardening and monitoring is the most effective at suppression of the USB-only (7.8% → 2.5%), but does not suppress phishing-only (22.8%). In turn, USB controls do not offer the same range of protection as awareness campaigns, however, they cut the physical-device route, which many blended campaigns hinge on, sharply.

The hybrid defense gives the best profile of all the measures. The total compromise decreases to 17.9%, which is more than 50% as compared to the baseline, and the detection rate is 35.7 %. The phishing feature dominates the combo composition in the hybrid setup (9.8%), with attacks that are USB only nearly annihilated (1.2%), and the mixed combo held (6.9%). This trend can be described as complementary: user-centered solutions decrease phishing vulnerabilities; device-based solutions decrease removable-media threat; and anomaly-based analytics increase the early signal detection and association.

These are supported by temporal dynamics. The time-to-compromise distributions of the baseline and hybrid conditions are overlaid in Figure 9, and this shows the strong rightward shift in the time-to-compromise distributions under hybrid conditions.

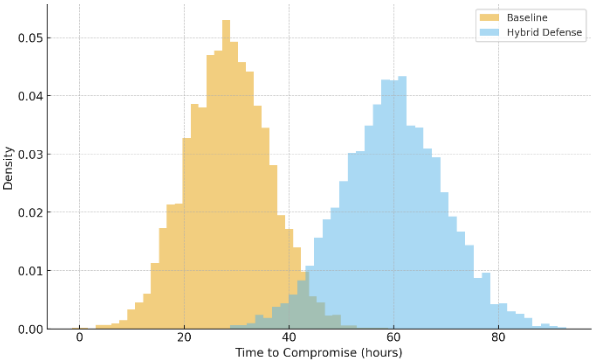

Fig. 9: Time-to-Compromise: Baseline vs. Hybrid Defense

In hybrid defenses, it is estimated that the median time to compromise would grow by a factor of 4, reaching about 60 hours instead of about 28.5 hours and that early compromises (less than 12 hours) would be considerably less common. A long dwell time before compromise enhances the probability of monitoring and response functions detecting the attacks before the exfiltration or privilege escalation. Combined, these findings indicate that the multiplicity rather than the additive nature of layered controls. The hybrid strategy provides proportionally better resilience to multivector, human-centric attacks by reducing the success in different pathways and increasing the detection and response window, as compared to individual measures.

4.3 Quantifying Vector Synergy

Synergy between vectors was assessed by contrasting observed multivector outcomes with an independence baseline. Let pp denote the phishing-only success probability, p_u the USB-only success probability, and pc the probability of compromise via either or both vectors. Under statistical independence, the expected union success is:

$$p_{ind} = p_p + p_u - (p_p \times p_u).$$

In the baseline configuration, $p_p = 0.245$ and $p_u = 0.078$, yielding $p_{ind} = 0.245 + 0.078 - (0.245 \times 0.078) = 0.305$. The observed overall success was $p_c = 0.387$, producing an excess margin of 0.387 - 0.305 = 0.082 (8.2 percentage points). This uplift indicates that coordinated, metadata-guided orchestration increases the probability of compromise beyond what independent operation of vectors would predict. Under a hybrid defense regime, $p_p = 0.098$ and $p_u = 0.012$ imply $p_{ind} = 0.109$, whereas the observed $p_c = 0.179$.

The resulting synergy margin is approximately 0.179 - 0.109 = 0.070 (7 percentage points).

Thus, even when layered controls are present, coordinated blending of vectors continues to confer a substantial advantage, on the order of 7–8 percentage points. To corroborate this effect, a logistic regression was estimated on trial-level outcomes with predictors including user risk score, defense scenario, a vector-combination indicator, and execution timing. The interaction term for phishing × USB was positive and statistically significant ($p < 0.01$), indicating that joint deployment elevates success beyond additive contributions. Collectively, these results show that treating vectors in isolation underestimates campaign efficacy; synergy remains a critical driver of breach likelihood across defensive postures.

4.4 Sensitivity and Robustness Analysis

The strength of the model was tested by manipulating the key parameters by a factor of ±50 percent and tracking resultant changes of the campaign. When the probability of USB plug-in was altered about a 10 per cent base value, significant changes in the yield of vectors were obtained: at 5 per cent USB-only success decreased by about half and the overall compromise rate was brought to about 34 per cent. USB-only success grew by about 15 percent, bringing the total compromise to approximately 4344 percent. The sensitivity confirms empirical research that reports significantly greater real-world plug-in behaviour in specific situations (e.g., 4598%), and highlights the behavioural reliance of the physical vector [3].

The phishing rate of clicking had the strongest effect. Comparing a 3% baseline to a 1.5% reduced the phishing-only success rate to 0.12 and the total compromise rate to 0.28-50, and to 0.36 and 48-50 respectively under an increased rate of 4.5%. These results confirm the superiority of click-resistance as the lever of control: the modestly obtained enhancements with the help of awareness training, secure-by-design interfaces, and content-filtering lead to disproportionately high risk-reductions. Reducing the detection latency by half (e.g., 48 h 2 48^{-} 24^{-}) reduced the total compromise by an absolute difference of 4 or so percentage points, which suggests that faster triage and alert processing has a meaningful impact on reducing attacker dwell time without additional changes. Conversely, reversing the order of launching the vector (phishing before USB or vice versa) had very little impact (<120pp), indicating that orchestration and coverage have a stronger impact than timeliness. Through parameter sweeps, the relative behavior of the contributions changed the same way, with phishing > USB, and combined-vector synergy ranking highest. Hybrid defences were always the least compromising and the ones that took the longest time to achieve intrusion success.

4.5 External Validity: Real-World Cases & Scholarly Evidence

In order to prove the reality of our simulation and the scale of the relative trends, we compare it to the described cases and empirical research.

4.5.1 USB Drop Tests and Practice Attacks.

A typical lab experiment entailed the testing of 297 USB flash drives on a university campus with the observed plug-in rates of between 45 percent and 98 percent and median connection time of about six and nine tenths minutes [3]. These data suggest that the human vulnerability to these vectors is notably greater than the conservative 10% assumption that was used in our model. Had the model assumed a 45 per cent plug-in rate, the probability of a successful attack using the USB would rise exponentially, suggesting that our simulated model of the USB devices is a lower bound. Correlated field surveys support the presence of the curiosity-induced USB usage in users. As an example, the report by Elie Bursztein entitled “Does Dropping USB Keys Really Work?,” presented at Black hat recorded plug-in rates of about 49 percent in similar campus experiments [39]. These empirical rates of plug-ins give credence to the possibility of USB-based attacks even among potentially informed populations.

USB breaches are still being reported in the wild. A recent case study published in Heliyon (a forensics journal) about an insider infiltrator who stole files via USB drives over an extended period of time also showed that the insider inserted overwritten USB sticks to avoid being caught [40]. This example underscores the ability of USB based vectors to work continuously and silently, which is consistent with the assumption of our simulation that detection is commonly delayed. Industrial reports also show that 79 per cent of threats posed by USB devices may result in the most devastating disruption [41], thus proving that the USB format is still very consequential in the operating environment. In addition, the exploits of BadUSB-type, which consist of USB-based devices pretending to be legitimate peripherals or injection of keystrokes, have been shown to be a real threat (see, e.g., off-path injection vulnerabilities reported by Dumitru et al. [42]). Although our simulation presupposed the injection of keylogger payloads, the actual USB attacks can be more insidious and powerful, thus increasing the actual rates of vectors above the simulated baseline.

4.5.2 AI-Driven Phishing Studies

With regard to phishing, the recent empirical studies show that spear-phishing messages constructed by artificial intelligence are just as effective as the spear-phishing messages that are carefully designed by the human professionals. AI-generated employment emails were in line with human-constructed emails, with 54 percent being clicked through, which is comparable to expert-created emails and significantly higher than the generic phishing emails (54 percent) [43]. Such results suggest that rivals armed with massive language models will always generate very misleading baits. The simulation used a more pessimistic click-through rate (3 per cent); nevertheless, AI-based phishing can go above this rate in a targeted campaign, signifying that the preeminence of the phishing vector in the simulation is moderate as compared to the real adversarial capacity.

Moreover, it has been noted that phishing messages generated by artificial intelligence are more effective at bypassing commercial and institutional spam filters, due to finer language nuances, which can blur automated fraud detection algorithms [44]. Ongoing studies on phishing recognition indicate that human subjects still demonstrate a significant degree of vulnerability, as most of them misclassify or ignore phishing warning signs, even in controlled experimental settings [45]. These empirical observations form the assumptions of the simulation of non-negligible underlying susceptibility. Aggregatively, the empirical USB-phishing statistics support the relative hierarchy of vectors used in simulation phishing dominance, USB significance, and combined vector dominance, and justify the space of underlying assumptions. The real-world compromise rates can be even higher than the calculated ones in settings with higher user error rates and more advanced USB attacks, which once again supports the idea that the given data are underestimates and not exaggerations.

5. DISCUSSION AND RECOMMENDATIONS

The findings of this paper indicate the perennial and dynamic importance of human factors as a key weakness in cyber security. The simulated attacks have shown that those that exploit human behaviour (through phishing, physical vectors such as USB devices and metadata-based profiling) remain highly effective despite the implementation of advanced technology controls. This complementary effect of these attack vectors makes them more powerful since each of them focuses on different aspects of human mental processes and business process. It was determined that despite the presence of defended layers, the likelihood of compromise is not trivial, which demonstrates the need to provide cyber security resilience with behavioural, technical, and policy-based interventions.

5.1 Discussion

The simulated campaigns provide three interconnected insights into the exploitation of the human factors, including the pre-eminence of phishing as entry-vector, the complementary nature of USB-based physical exploits, and the strategic benefit of synergy of vectors. Phishing is the most effective vehicle since it directly plays around with indicators of trust and authority to affect fast, or even subconscious, decision-making [6]. The results that even such small differences in click-through rates have imbalanced impacts on overall compromise (±20 percentage points) are supporting evidence of past empirical studies on user behaviour as the key predictor of breach probability [2, 11].

USB-based vectors are included to expand the knowledge of physical-layer risks which are often ignored in modern digital security models. Although numerous system organizations focus on network firewalls and patching software, it has been proven that the hardware-based exploits bypass the digital controls completely after breach of physical trust gates. The results of the simulation, which show USB-only attacks of around 7.8% during baseline, and more than 50% variance with changes in behaviour, are comparable to empirical plug-ins of 45 98 per cent [3]. These results confirm real life experience where curiosity, altruism or work related habits cause employees to operate with unknown devices thus showing that physical trust is still entrenched in organisational culture.

One of the most important lessons can be viewed in the 7 to 8 percent synergy margin that is being seen in all the scenarios: in case attackers use vectors in a coordinated manner, the general success rates will be higher than in cases where individual probabilities are used. This interaction effect is consistent with literature on blended threats that highlights the existence of cognitive overload and cross-channel confusion in hybrid campaigns that reduce efficacy in defense [15]. The simulated effect of synergy also coincides with case studies of advanced persistent threat (APT) whereby layered tactics dominate segmented defense, including the FIN7 and Stuxnet operations. Metadata profiling is used to target precisely in such campaigns, USB exploits allow building persistence, and phishing can be used to fast-track initial access.

The analysis also confirms the fact that defense effectiveness is multiplicative and not additive. Compromises were reduced by over 50%, and the median time-to-compromise was changed to 60 hours from 28.5 hours with the hybrid defence, which included awareness training, USB port control, and anomaly-based monitoring. These enhancements explain the role of various defensive layers that cover the blind spots of each other. Technical hardening will remove routes of physical access, but awareness training will modify behavioural probabilities and analytics will reduce the detection latency. The combination of these results in a robust positive feedback between human caution and machine-aided detection [18]. Nevertheless, there is still residual compromise (~17.9) under hybrid defences, which highlights that in socio-technical systems requiring human decision-making, total prevention is not possible.

Another aspect of debate is the time-dynamics of compromise. The histogram and survival analysis have shown that the majority successful breaches are in the 24-48 hours period after launching a campaign, which highlights a close detection timeline. This observation is consistent with the incident-response reports that formulated that mean time to detect (MTTD) exceeds 72 hours in most organisations [13]. A small decrease in latency detection generated statistically significant gains in the simulation results (~4 percentage points). Thus, the improvement of real-time analytics and fast-response processes is a real-life defensive benefit that does not depend on being trained or having infrastructure in place.

The parameter sweeps in the simulation prove that the qualitative naming of the impact of vectors (phishing > USB > combined synergy) is consistent. Such consistency supports the external validity of the results compared to empirical research. Indicatively, Heiding et al. [43] detected that AI-generated phishing has an average of more than 50 percent click rates, which the inference made in the simulation showed that phishing was the strongest human exploit type. In the same way, the supposedly permanent risks of human curiosity and physical accessibility are supported by field evidence of USB drop success [39]. The agreement between the simulated and the real results further increases the credibility of the model as a conceptual proxy to organisational threat analysis.

The discussion also explains important implications on organisational policy. To begin with, security awareness training should not be limited to any of the fixed or compliance-based models; simulated outcomes show that even slight changes in behaviour may have an enormous impact on the rates of compromising. As a result, periodic modules should be replaced by adaptive, continuous, and gamified training modalities. Second, the endpoint management should include USB hardening and device control policies. It has been established through empirical research that banning unregistered external drives and the use of automatic encryption can reduce incidents involving removable media significantly [23]. Third, metadata protection that is frequently neglected by cybersecurity frameworks should be prioritised urgently. The growing complexity of metadata profiling requires organisational policies limiting the publicity of data (e.g. employee schedule, contacts list) and metadata scrubbing of documents and images uploaded on the web [7].

The aspect of psychological resilience in cybersecurity can also be highlighted by the fact that compromise has continued to be witnessed in various situations in the defence field. Such manipulation is impossible to address entirely using technical measures, which make use of the cognitive biases of authority, urgency, and familiarity [21]. It is therefore upon organisations to go beyond awareness training to form a culture of scepticism and considered thought under pressure. The principles of human-centric design (including security nudges, real-time feedback in user interfaces, or contextual prompts preceding the implementation of risky behaviour) can direct behaviour towards making safer decisions without necessarily having to rely on memory or training.

Lastly, the intersection of AI-based fraud and human fallacy poses ethical and regulatory issues of great concern. The fact that the linguistic and emotional clues are reproducible by the generative models on the scale threatens the integrity of identity and undermines the trust in a digital communication. With the opponents becoming more automated in their persuasion methods, the traditional methods used to prevent these are becoming ineffective by the day, including awareness campaigns and signature-based filtering systems. Emergent landscape thus requires regulatory innovation such as the watermarking of AI-generated content and investment in counter-AI infrastructures that have the ability to recognize synthetic text and media.

5.2 Recommendations

Resting on the empirical evidence and simulation insights provided in the current paper, the following strategic suggestions are hypothesised to increase the organisational resilience to human-element multivector attacks.

- **Institute Multi-layered human-centric defense designs:** Security architectures should be built with lessons of human behaviour and technical protection. Hybrid defences must have three mutually enforcing pillars namely: (i) continuous awareness programs aimed at making them less susceptible; (ii) endpoint and peripheral controls that prevent unauthorised USB access; and (iii) anomaly-based monitoring systems that can recognise abnormal behavioural patterns. These layers need to work with standard dashboards whereby an alert by one of the vectors (such as suspicious USB insertion) initiates contextual verification procedures in other vectors (such as simultaneous phishing alerts).
- **Embrace Behavioural Analytics and Cognitive Modelling to Early Warning:** The classical intrusion detection systems should develop to include behavioural user telemetry including response time to emails, abnormalities in logs, and abnormalities in the use of devices. Machine-learning algorithms have the potential to represent baseline cognitive-behavioural rhythms thus indicating variation to be reviewed. This method, which could also be called cognitive intrusion detection, is consistent with the current body of research that recommends adaptive security models that react to human error tendencies.
- **Continuous and Contextual Awareness Training should be Institutionalised:** A single training event cannot work in the age of AI-driven deception. Training programs can and must be contextual, iterative, data-based, with real-time phishing simulations, role-based training, and behavioural feedback. Empirical research shows that the repeated micro-learning interventions may decrease the number of the click-through by up to 40. The training should also be based on the affective aspects of persuasion, as they will allow the employees to discover the urgency and authority signals that are regularly used in AI-generated phishing attacks.
- **Enhance Physical and Peripheral Controls:** Organisations should tighten control on the use of external media by enforcing strict rules on its use to enhance physical and peripheral security. Such measures include disabling USB autorun, device whitelisting, and encrypted corporate drives with endpoint validation. It is confirmed by empirical results that even the simplest port-control interventions can decrease USB-based infections by over seventy percent. At the same time, the social-engineering aspects of physical baiting should be addressed in the awareness campaign, educating the personnel to report the suspicious equipment instead of engaging with it.
- **AI-Detection and Digital Provenance Tools Could be integrated:** Countermeasures to weaponisation of AI models in phishing and impersonation should include machine-learning-based text, audio and video synthetic-text detectors. Such tools as deep-fake detectors, email provenance verification systems (e.g., DMARC, DKIM) must be institutionalised. Regulatory level: Organisations should be the proponents of regulatory mechanisms to impose watermarking or traceability of AI-generated content to enable the process of attribution and responsibility.
- **Minimise Detection Latency by Automation:** The latency of detection turned out as a key determinant in the simulation. Priorities should be given to automated incident response and real-time correlation of user, device and network data. Security orchestration and response (SOAR) platforms can help to reduce response times and provide pre-emptive isolation of impacted systems. These steps are operationalizations of the simulation result that compromise is minimized by about four percentage points by decreasing detection latency by half.
- **Enhance Continuous Investigations on Human-AI Interaction in Security Scenarios:** Human cognition and AI-related deception is an area where there is limited research on the interaction. Further studies are needed to explore the changes in cognitive biases in response to the use of synthetic content and ways to adapt defensive AI systems to respond to manipulative behavior. The ability

to design the following generation of socio-technical safeguards will require field experiments combining behavioural psychology, human-computer interaction, and cybersecurity analytics.

6. CONCLUSION

The current paper presented three key vectors of attack based on the human element, including USB-based attacks, AI-driven phishing, and metadata-based surveillance, and assessed the synergistic enhancement of cyber risk that is created by the intersection of these three vectors. An effective approach to a balanced mix of theoretical synthesis in the form of peer-reviewed literature with simulated multivector attack situations was adopted. The results clearly show that when both behavioural vulnerabilities of human actors are combined with technological weaknesses, traditional defensive architecture cannot provide appropriate protection. In this discussion, it is highlighted that cybersecurity cannot be regarded as a field of technical study, but it is a socio-technical system that essentially relies on the cognitive and cultural behaviours of humans, as well as, organizational behaviour.

According to the empirical evidence, phishing has remained the most common and effective attacker form, as it has the ability to manipulate the human trust, authority, and a sense of urgency. The probability of compromise in response to even marginal changes in the click behaviour of users is significant, which highlights the primacy of psychological factors in cybersecurity. Physical exploits via USB, although apparently low-tech, still represent a major threat, mainly due to the exploitation of human curiosity and the daily operation routines. Metadata surveillance proves that a malicious actor may breach privacy and security without the use of malware and reveal the structural weaknesses of digital transparency and data management. These vectors when combined create a synergistic amplification effect which enhances the overall success probability of an attack by over seven to eight percentage points instead of independent expectations. The synergy is a reflection of the threat environment of the present day where physical, social, and informational space is actively used to enable swift and covert intrusions.

The defense exercises also clarified that the convergence of technical, behavioural and analytical layers of defense, referred to as hybrid defence strategy, is the only defence approach that provides the greatest minimization of risk. Vulnerability to social engineering was reduced due to awareness training, USB port controls ensured that devices could not be used without authorization, and anomaly-based analytics decreased the detection latency. All these actions prolonged time-to-compromise and enhanced detection rates. Nevertheless, despite the successful hybrid designs, a compromise likelihood of about 18% remained, which supports the idea that no technical system can leave a hundred percent of human-induced vulnerabilities. As such, cybersecurity strategy should be built on resilience and not prevention.

More importantly, the study confirmed that human-oriented security culture and cognitive preparedness are just as vital as investment in technology. Scenario-based training, continuous learning, and transparency should substitute reactive and compliance-based strategies in organisations. The results also accentuate the increased difficulty of artificial intelligence in enhancing deception, as now generative models can mimic people, create messages that are contextually relevant, and circumvent conventional censorware. This further development requires its own dynamically adaptive defenses such as AI-based provenance of content, digital watermarking, and dynamic threat intelligence.

Finally, the research adds to the body of academic literature and real-life cybersecurity management practices by showing that the human factor is the most significant weak point and the strongest defense tool. The existence of a robust cybersecurity ecosystem is that the human aspect should be incorporated in the creation of the ecosystem, fostering awareness, scepticism, and team-work at all levels of the organization. Since the attackers exploit the intersection of the digital and human vectors, defenders need to counteract them with holistic socio-technical systems that make people, processes, and technologies work together. Finally, the only way to ensure digital trust in the age of AI-based dishonesty is to change all the users into both a potential threat and a knowledgeable and active agent in terms of protection.

This broad insight defines cybersecurity not as a technological fight against intrusion, but rather a marriage of human intuition and opponent manipulation, which can only be won by constantly evolving, collaborating, and intelligently combining human and machine abilities.

7. ACKNOWLEDGMENTS

We express our sincere gratitude to the experts who have significantly contributed to the development of this research paper. Their insights, guidance, and support were invaluable in shaping this work. We want to acknowledge the contributions of the following authors, particularly:

Allan Munyira, Yeshiva University - Cybersecurity (Email: munyiraallan@gmail.com), for leading the research and manuscript preparation.

Carrol Donna Kudaro, Yeshiva University - Cybersecurity (Email: carrolkudaro@gmail.com), for her critical analysis and valuable inputs throughout the study.

Collins Katende, Yeshiva University - Cybersecurity (Email: katendecollinz@gmail.com), for his support in data interpretation and framework validation.

Hamuza Senyonga, Yeshiva University - Cybersecurity (Email: senyonga.hamuza@gmail.com), for his technical expertise and assistance in data analysis.

We also appreciate the collaborative spirit and commitment shown by all contributors, whose collective efforts made this research possible.

8. REFERENCES

[1] Jabir, R., J. Le, and C. Nguyen. 2025. “Phishing Attacks in the Age of Generative Artificial Intelligence: A Systematic Review of Human Factors.” *AI* 6, no. 8 (July 31): 174. https://www.mdpi.com/2673-2688/6/8/174.

[2] Khadka, K., and A. B. Ullah. 2025. “Human factors in cybersecurity: an interdisciplinary review and framework proposal.” *Int J Inf Secur* 24, no. 3 (June 29): 119. https://link.springer.com/10.1007/s10207-025-01032-0.

[3] Tischer, M., Z. Durumeric, S. Foster, S. Duan, A. Mori, E. Bursztein, et al. 2016. “Users Really Do Plug in USB Drives They Find.” In *Proceedings—2016 IEEE Symposium on Security and Privacy, SP 2016*, 306–19. Accessed October 2, 2025.

https://blog.knowbe4.com/users-really-do-plug-in-usb-drives-they-find.

[4] Rohan, R., S. Funilkul, D. Pal, and W. Chutimaskul. 2021. "Understanding of Human Factors in Cybersecurity: A Systematic Literature Review." In *2021 International Conference on Computational Performance Evaluation (ComPE)*, 133–40. IEEE. https://ieeexplore.ieee.org/document/9752358/.

[5] Department of Financial Services. 2020. "Twitter Investigation Report | Department of Financial Services." New York State. Accessed September 29, 2025. https://www.dfs.ny.gov/Twitter_Report.

[6] Webb, T., K. J. Holyoak, and H. Lu. 2023. "Emergent analogical reasoning in large language models." *Nat Hum Behav* 7, no. 9 (July 31): 1526–41. https://www.nature.com/articles/s41562-023-01659-w.

[7] Ravindranath, R. 2024. "Security and Privacy Considerations of Metadat." *Int J Comput Tech* 11, no. 6: 1–5.

[8] Al-Sarawi, S., M. Anbar, R. Abdullah, and A. B. Al Hawari. 2020. "Internet of Things Market Analysis Forecasts, 2020–2030." In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, 449–53. IEEE. https://ieeexplore.ieee.org/document/9210375/.

[9] Sicari, S., A. Rizzardi, L. A. Grieco, and A. Coen-Porisini. 2015. "Security, privacy and trust in Internet of Things: The road ahead." *Comput Networks* 76 (January): 146–64. https://linkinghub.elsevier.com/retrieve/pii/S1389128614003971.

[10] Roman, R., J. Zhou, and J. Lopez. 2013. "On the features and challenges of security and privacy in distributed internet of things." *Comput Networks* 57, no. 10 (July): 2266–79. https://linkinghub.elsevier.com/retrieve/pii/S1389128613000054.

[11] Antonakakis, M., T. April, M. Bernhard, E. Bursztein, Z. Durumeric, A. J. Halderman, et al. 2016. "Understanding the Mirai Botnet." In *Proceedings of the 26th USENIX Security Symposium*, 72. Accessed October 6, 2025. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis.

[12] Nicho, M., and I. Sabry. 2023. "Bypassing Multiple Security Layers Using Malicious USB Human Interface Device." In *Proceedings of the 9th International Conference on Information Systems Security and Privacy*, 501–8. SCITEPRESS - Science and Technology Publications. https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0011677100003405.

[13] Verizon. 2025. *2025 Data Breach Investigations Report*. Accessed September 29, 2025. https://www.verizon.com/business/resources/reports/dbir/.

[14] Herrmann, D., and H. Federrath. 2017. "Editorial: 30th IFIP International Information Security Conference (IFIP SEC 2015)." *Comput Secur* 67 (June): 266. https://linkinghub.elsevier.com/retrieve/pii/S016740481730069X.

[15] Sen, Ö., B. Ivanov, C. Kloos, C. Zöll, P. Lutat, M. Henze, et al. 2025. "Simulation of multi-stage attack and defense mechanisms in smart grids." *Int J Crit Infrastruct Prot* 48 (March): 100727. https://linkinghub.elsevier.com/retrieve/pii/S1874548224000684.

[16] Cloudflare. n.d. "Risk grows as multi-vector attacks become the norm." theNET. Accessed October 2, 2025. https://www.cloudflare.com/the-net/multi-vector-threats/.

[17] Baker, E., and M. Cartier. 2025. "Phishing Trends Report (Updated for 2025)." Hoxhunt. Accessed September 29, 2025. https://hoxhunt.com/guide/phishing-trends-report.

[18] García-Teodoro, P., J. Díaz-Verdejo, G. Maciá-Fernández, and E. Vázquez. 2009. "Anomaly-based network intrusion detection: Techniques, systems and challenges." *Comput Secur* 28, no. 1–2 (February): 18–28. https://linkinghub.elsevier.com/retrieve/pii/S0167404808000692.

[19] Axelsson, S. 2000. "The base-rate fallacy and the difficulty of intrusion detection." *ACM Trans Inf Syst Secur* 3, no. 3 (August): 186–205. https://dl.acm.org/doi/10.1145/357830.357849.

[20] Pham, H. C., D. D. Pham, L. Brennan, and J. Richardson. 2017. "Information Security and People: A Conundrum for Compliance." *Australas J Inf Syst* 21 (January 18). https://ajis.aaisnet.org/index.php/ajis/article/view/1321.

[21] Reed, S., K. Zolna, E. Parisotto, S. G. Colmenarejo, A. Novikov, G. Barth-Maron, et al. 2022. "A Generalist Agent." November 11. arXiv. http://arxiv.org/abs/2205.06175.

[22] Wahanani, H., M. Idhom, and D. R. Kurniawan. 2020. "Exploit remote attack test in operating system using arduino micro." *J Phys Conf Ser* 1569 (July): 022038. https://iopscience.iop.org/article/10.1088/1742-6596/1569/2/022038.

[23] Nicho, M., and I. Sabry. 2023. "Bypassing Multiple Security Layers Using Malicious USB Human Interface Device." In *Proceedings of the 9th International Conference on Information Systems Security and Privacy*, 501–8. SCITEPRESS - Science and Technology Publications. https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0011677100003405.

[24] Nasution, S. M., Y. Purwanto, A. Virgono, and M. R. Y. Tambunan. 2015. "Integration of autonomous sender for hidden log data on kleptoware for supporting physical penetration testing." In *2015 1st International Conference on Wireless and Telematics (ICWT)*, 1–5. IEEE. http://ieeexplore.ieee.org/document/7449205/.

[25] Singh, D., A. K. Biswal, D. Samanta, D. Singh, and H. N. Lee. 2022. "Juice Jacking: Security Issues and Improvements in USB Technology." *Sustainability* 14, no. 2 (January 14): 939. https://www.mdpi.com/2071-1050/14/2/939.

[26] Cronin, P., X. Gao, H. Wang, and C. Cotton. 2022. "Time-Print: Authenticating USB Flash Drives with Novel Timing Fingerprints." In *2022 IEEE Symposium on Security and Privacy (SP)*, 1002–17. IEEE. https://ieeexplore.ieee.org/document/9833595/.

[27] Joven, R., and N. K. Choon. 2023. "The Spies Who Loved You: Infected USB Drives to Steal Secrets." Google Cloud. Accessed October 2, 2025. https://cloud.google.com/blog/topics/threat-intelligence/infected-usb-steal-secrets/.

[28] Nissim, N., R. Yahalom, and Y. Elovici. 2017. “USB-based attacks.” *Comput Secur* 70 (September): 675–88. https://linkinghub.elsevier.com/retrieve/pii/S0167404817301578.

[29] Zhang, R., A. Bello, and J. L. Foster. 2023. “BYOD Security: Using Dual Process Theory to Adapt Effective Security Habits in BYOD.” In *Proceedings of the Future Technologies Conference (FTC) 2022, Volume 2 FTC 2022 2022 Lecture Notes in Networks and Systems*, 372–86. Springer. https://link.springer.com/10.1007/978-3-031-18458-1_26.

[30] Neupane, S., I. A. Fernandez, S. Mittal, and S. Rahimi. 2023. “Impacts and Risk of Generative AI Technology on Cyber Defense.” June 22. arXiv. http://arxiv.org/abs/2306.13033.

[31] Kolobrodova, A. 2024. “Metadata 102 — What is communications metadata and why do we care about it?” Accessed October 2, 2025. https://freedom.press/digisec/blog/metadata-102/.

[32] Office of the Chief Information Security Officer. 2020. “The FIN7 Cyber Actors Targeting US Businesses through USB Keystroke Injection Attacks |.” OCISO. Accessed October 3, 2025. https://ociso.ucla.edu/news/fin7-cyber-actors-targeting-us-businesses-through-usb-keystroke-injection-attacks?utm_source=https://www.google.com/search?q=chatgpt.com.

[33] Francia, J., D. Hansen, B. Schooley, M. Taylor, S. Murray, and G. Snow. 2025. “Assessing AI vs Human-Authored Spear Phishing SMS Attacks: An Empirical Study.” March 19. arXiv. http://arxiv.org/abs/2406.13049.

[34] Perez, B., M. Musolesi, and G. Stringhini. 2018. “You are your Metadata: Identification and Obfuscation of Social Media Users using Metadata Information.” May 14. arXiv. http://arxiv.org/abs/1803.10133.

[35] Fakiha, B. S. 2024. “Forensic analysis of bad USB attacks: A methodology for detecting and mitigating malicious USB device activities.” *Edelweiss Appl Sci Technol* 8, no. 5 (September 19): 1090–100. https://learning-gate.com/index.php/2576-8484/article/view/1809.

[36] Li, W., S. Manickam, Y. W. Chong, Y. He, H. Y. Li, and B. Li. 2025. “ByteBait USB: a robust simulation toolkit for badUSB phishing campaign.” *J King Saud Univ Comput Inf Sci* 37, no. 5 (July 1): 91. https://link.springer.com/10.1007/s44443-025-00067-6.

[37] Alkhalil, Z., C. Hewage, L. Nawaf, and I. Khan. 2021. “Phishing Attacks: A Recent Comprehensive Study and a New Anatomy.” *Front Comput Sci* 3 (March 9). https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060/full.

[38] Corona, I., G. Giacinto, C. Mazzariello, F. Roli, and C. Sansone. 2009. “Information fusion for computer security: State of the art and open issues.” *Inf Fusion* 10, no. 4 (October): 274–84. https://linkinghub.elsevier.com/retrieve/pii/S156625350900030X.

[39] Bursztein, E. 2016. “Does dropping usb drives really work?” Black Hat. Accessed October 4, 2025. https://ly.tl/malusb.

[40] Chung, M. H. (Miles), Y. Yang (Alisha), L. Wang, G. Cento, K. Jerath, P. Taank, et al. 2023. “Enhancing cybersecurity situation awareness through visualization: A USB data exfiltration case study.” *Heliyon* 9, no. 1 (January 1): e13025. https://doi.org/10.1016/j.heliyon.2023.e13025.

[41] Ribeiro, A. 2021. “USB removable media still acts as an initial attack vector in OT environments.” Industrial Cyber. Accessed October 4, 2025. https://industrialcyber.co/threats-attacks/usb-removable-media-still-acts-as-an-initial-attack-vector-in-ot-environments/?utm_source=https://www.google.com/search?q=chatgpt.com.

[42] Dumitru, R., D. Genkin, A. Wabnitz, and Y. Yarom. 2022. “The Impostor Among US(B): Off-Path Injection Attacks on USB Communications.” November 3. arXiv. http://arxiv.org/abs/2211.01109.

[43] Heiding, F., S. Lermen, A. Kao, B. Schneier, and A. Vishwanath. 2024. “Evaluating Large Language Models’ Capability to Launch Fully Automated Spear Phishing Campaigns: Validated on Human Subjects.” November 30. arXiv. http://arxiv.org/abs/2412.00586.

[44] Shreyas Kumar, Anisha Menezes, Sarthak Giri, and Srujan Kotikela. 2024. “What The Phish! Effects of AI on Phishing Attacks and Defense.” *Int Conf AI Res* 4, no. 1 (December 4): 218–26. https://papers.academic-conferences.org/index.php/icair/article/view/3224.

[45] Carroll, F., J. A. Adejobi, and R. Montasari. 2022. “How Good Are We at Detecting a Phishing Attack? Investigating the Evolving Phishing Attack Email and Why It Continues to Successfully Deceive Society.” *SN Comput Sci* 3, no. 2 (March 23): 170. https://link.springer.com/10.1007/s42979-022-01069-1.

[46] Oner, U., O. Cetin, and E. Savas. 2025. “Human factors in phishing: Understanding susceptibility and resilience.” *Comput Stand Interfaces* 94 (August): 104014. https://linkinghub.elsevier.com/retrieve/pii/S0920548925000431.