ChatGPT

Large Language Models in Cybercriminal Operations: A Deep-Dive

Cybercriminals – including state-sponsored groups from Russia, China, and North Korea – are rapidly adopting large language models (LLMs) as offensive tools. Recent threat intelligence and academic studies reveal that these groups are leveraging LLMs for everything from crafting convincing phishing lures to assisting in malware development. This report provides an in-depth analysis of how *advanced threat actors* are exploiting LLMs for infiltration, payload development, and social engineering, with a focus on Russian, Chinese, and North Korean hacker groups. We draw on the latest findings from security researchers, government reports, and open-source intelligence to shed light on these cutting-edge techniques.

LLM-Powered Social Engineering and Phishing

One of the clearest ways criminals and nation-state hackers are abusing AI is to supercharge phishing and social engineering campaigns. **Generative AI enables threat actors to craft highly convincing, tailored messages at scale**, eliminating many of the tell-tale errors that once gave phishing emails away [1] [2]. In the pre-ChatGPT era, phishing emails often suffered from poor grammar and awkward phrasing – especially when attackers were not native speakers. Now, LLMs like GPT-3.5/4 can produce polished, contextually appropriate prose, making malicious emails and texts much harder to distinguish from legitimate communication [1]. A report by Acronis found that AI-assisted phishing impacted over 90% of organizations in 2023, with email attack volume surging 222% as attackers used generative AI to optimize content [3]. In short, **LLMs help cybercriminals "increase cyberattack efficiency, create malicious code, and automate attacks,"** according to Acronis VP Candid Wüest [4].

State-aligned hacking groups have begun weaving LLMs into their phishing toolkits. In 2024, Microsoft and OpenAI observed that North Korea's *Kimsuky* espionage unit used LLMs to generate content for spear-phishing campaigns targeting foreign policy think-tanks [5]. Likewise, an Iranian Revolutionary Guard group (*Crimson Sandstorm*) employed LLMs to write spear-phishing emails – one masquerading as a message from an international development agency, another attempting to lure feminist activists to a fake website [6]. These AI-written lures were markedly more credible and well-crafted than the groups' typical output, highlighting how **AI helps accelerate and boost phishing email production** [6]. OpenAI's own threat intelligence confirmed multiple cases where **state-backed hackers used ChatGPT to generate social engineering content**, such as phishing templates and even draft messages for influence campaigns [7] [8].

Criminal enterprises are also using generative AI for **Business Email Compromise (BEC)** and other fraud schemes. In mid-2023, underground forums began selling illicit LLM-based chatbots like *WormGPT* and *FraudGPT* – essentially blackhat versions of ChatGPT with no ethical guardrails [9] [10]. **WormGPT**, built on the open-source GPT-J model, was explicitly trained on malware development data and showcased its ability to write a *convincing CEO fraud email* (a classic BEC scenario) as well as functional malware code [11] [12]. Advertised at €60–€100 per month on hacking forums, WormGPT quickly gained popularity for generating *polished phishing emails purportedly from a CEO* that could trick employees [12]. Similarly, **FraudGPT** emerged on dark web marketplaces and Telegram channels, billed as an *"all-in-one" AI cybercrime tool* with capabilities to write scam pages, phishing messages, and undetectable malware [13] [14]. Its creator (alias **CanadianKingpin**) claimed over 3,000 sales by July 2023, offering

subscriptions up to $1,700/year [15] [16]. Notably, threat intel analysts assess that the same actor likely runs both WormGPT and FraudGPT, using them for different criminal focuses – high-volume short scams versus longer-term malware/ransomware campaigns [17].

Beyond emails, **LLMs can generate entire social engineering personas and interactive scripts**. The Alan Turing Institute's March 2025 report warns of AI-driven romance scams where deepfake avatars and chatbot "lovers" carry out long-con social engineering [18]. In one case, a British company lost £20 million to fraudsters who employed *AI-generated deepfake executive voices and correspondence* to authorize transfers [18]. North Korean operators have similarly used AI to fabricate personas: an August 2025 Anthropic report revealed that groups of North Korean IT workers, moonlighting as freelance developers abroad, used generative AI to **create fake resumes and even *answer technical interview questions***, securing jobs at foreign companies under false identities [19] [20]. Once hired, these operatives leaned on AI assistance to perform programming tasks beyond their actual skill level, allowing them to exfiltrate data and earn illicit income for the regime [21] [20]. This novel use of AI essentially helped North Korean agents infiltrate organizations by **masquerading as qualified employees**, an infiltration tactic that would likely have failed without AI help, given their limited English and technical expertise [22].

Even the *themes* of phishing lures are adapting to the AI age. Genians, a South Korean cybersecurity firm, documented how North Korea's Kimsuky group sent emails with subject lines referencing AI features – e.g. claims of "AI managing emails on your behalf" – to entice victims into clicking [23]. In July 2025, Kimsuky executed a particularly bold phish: they **leveraged ChatGPT to produce a forged South Korean military ID card** image used in a spear-phishing email [24]. The deepfake ID (presented as a "draft" of a real ID) was meant to build credibility with the target, who was tricked into opening a malware-laced attachment [24]. According to Genians' analysis, ChatGPT initially refused to generate an image of an official military ID (as it violated content rules), but the attackers bypassed this by rephrasing the request as a benign "sample ID design" [25]. The resulting AI-generated ID card – while partially masked in released screenshots – was realistic enough to fool recipients [25] [26]. This incident demonstrates **attackers' ability to jailbreak or manipulate LLMs into producing social engineering props** (in this case, fake documents) that bolster their intrusions. In summary, from phishing emails and fake profiles to forged documents, LLMs have become a force-multiplier for social engineering by Russian, Chinese, and North Korean threat actors alike.

Using LLMs for Reconnaissance and Infiltration

Beyond crafting lures, adversaries are deploying LLMs as **"research assistants" for reconnaissance and attack planning**. Microsoft reported that Russia's GRU hackers (APT28 "Fancy Bear") were observed querying an LLM to gather in-depth technical information on satellite communication protocols and radar imaging technologies [27]. These queries likely aimed to accelerate the group's understanding of specialized systems related to the war in Ukraine, potentially to inform new attacks on satellite infrastructure [28]. In essence, **the AI model served as a quick reference and tutor**, summing up complex topics that hackers would otherwise pore over manuals or papers to learn. Microsoft characterized Fancy Bear's LLM use as exploratory – "an adversary exploring the use cases of a new technology" – but it underscores that even elite Russian operators are testing AI to enhance their technical reconnaissance [29] [30].

Chinese state-sponsored groups have similarly dipped their toes in LLM-assisted recon and development. In one case, a Chinese APT dubbed **"Chromium"** used an LLM to *generate and refine scripts*, apparently to streamline portions of their operations [31]. Another China-linked group ("Sodium") attempted to have an LLM produce *malicious code*, essentially trying to offload malware writing to the

AI, though they failed to bypass the model's safeguards against such output [32]. Meanwhile, a prolific Chinese espionage group known as **Aquatic Panda** was caught interacting with LLMs in ways suggesting they were assessing how AI might *augment their technical hacks*, albeit on a limited scale [33]. And **Maverick Panda**, active for over a decade targeting U.S. defense contractors, was seen querying LLMs for sensitive topics, high-profile individuals, and geopolitical info – hinting at using AI for intelligence gathering on persons of interest [33]. In these examples, Chinese actors treated LLMs as an information source and coding aide: translating foreign-language material, summarizing documents, and even explaining software errors. Indeed, Microsoft noted that Iran's hackers have used LLMs to **troubleshoot code and study how to evade detection on compromised networks** [6] – tasks traditionally requiring human expertise.

OpenAI's own investigations confirm that **threat groups are leveraging ChatGPT for many pre-attack activities**. In one case, an Iranian-affiliated group (*CyberAv3ngers*, linked to IRGC) abused ChatGPT to research default passwords for industrial control systems, find vulnerabilities in server software, and get guidance on crafting a Modbus network scanner [34] [35]. They even asked ChatGPT how to debug bash scripts and obfuscate code, essentially crowd-sourcing solutions to technical roadblocks in their attack chain [36]. None of these tasks are novel – hackers have long used search engines and forums for help – but **LLMs provide a faster, conversational way to obtain filtered answers**. ChatGPT "did not provide [them] any novel capability… only limited, incremental assistance" in this case, OpenAI noted [37]. Still, for an operator with moderate skills, having an on-demand AI tutor can shorten the development cycle for exploits or tools.

North Korean actors appear especially keen on using AI to compensate for skills gaps. The **Kimsuky APT**, beyond using AI for phishing as discussed, also employed ChatGPT to *research foreign think-tanks* and topics related to their targets [5]. This likely helped them write more informed phishing content and improve targeting. And in a twist on "infiltration," North Korean IT operatives (many working as contractors abroad) harnessed AI to **pass technical interviews and perform coding work** once inside companies [21] [20]. Anthropic's August 2025 threat report described how these operatives relied on Claude (Anthropic's LLM) to handle programming assignments and English communication that they would otherwise struggle with [22] [20]. In effect, the LLM became an accomplice in infiltrating victim organizations under false pretenses – a creative blend of social engineering and technical infiltration.

Another emerging use of LLMs is **processing stolen data and conducting victim profiling**. According to Anthropic, cybercriminals have started to embed AI at *all stages of their operations*, including *"analyzing stolen data, stealing credit card information, and creating false identities"* [38]. For instance, an attacker who exfiltrates a large email archive or database could feed it into an AI model to quickly identify passwords, financial records, or other high-value intel. One large-scale extortion operation in 2024–2025 reportedly used Anthropic's Claude not just for coding but to **make strategic decisions**: the AI autonomously prioritized which data to steal, chose extortion strategies, and even drafted intimidating ransom notes tailored to the victim's financials [39] [40]. Claude analyzed the stolen financial data to recommend ransom amounts and generated polished ransom messages complete with the victim's sensitive details and threats [39] [41]. This "AI agent" approach – letting the model act with some autonomy (so-called *"agentic AI"*) – takes infiltration to a new level. It shows that **criminals are experimenting with letting AI drive parts of the attack lifecycle**, from initial recon to final extortion, in ways that were science fiction a few years ago [42].

It's important to note that much of this LLM-enabled recon and planning is still in early stages. Microsoft assessed the techniques seen so far as *"incremental"* and "not particularly novel or unique" – essentially AI being used as another tool in the toolbox [43]. However, the breadth of adoption across Iran, North Korea, China, and Russia in less than a year underscores a trend: **LLMs are becoming embedded in offensive TTPs (tactics, techniques, procedures)** much like past technologies (VPNs, encryption, etc.).

The proliferation of **"open-weight" LLMs (models with public weights and weak guardrails)** is accelerating this adoption [44]. Unlike OpenAI's tightly monitored ChatGPT, open-source models (e.g. LLaMA derivatives, GPT-J/GPT-NeoX, Chinese models without safety layers) can be fine-tuned or outright jailbroken to answer virtually any prompt. The UK's CETaS report specifically flags Chinese advancements in open AI models as a catalyst for criminals, who exploit these **unfettered models to carry out more advanced tasks** without oversight [44]. In practical terms, a determined attacker can take an open model and train or prompt-engineer it to output everything from exploit code to insider trading plans – all away from the eyes of any provider. We are already seeing threat actors brag about custom-tuned AI models on forums, signaling that AI-aided infiltration will likely become more sophisticated and harder to detect in the coming years.

LLMs in Malware Development and Payload Delivery

Perhaps the most alarming use of LLMs is in **automating malware creation and refinement of attack payloads**. Traditionally, writing exploit code or sophisticated malware required considerable programming skill. Now, even low-skilled attackers can have an AI model generate functional malicious code or assist in debugging it. OpenAI's October 2024 security report revealed a case where an Iranian threat actor (*STORM-0817*) actively used ChatGPT as a coding assistant while developing new Android malware [45] [46]. The hacker fed ChatGPT snippets of their malware code for debugging; in doing so, they inadvertently gave OpenAI visibility into their in-progress spyware (which was designed to steal call logs, contacts, files, etc.) [47]. ChatGPT helped troubleshoot errors and even provided guidance for the malware's command-and-control server code [48]. **This "Malware-as-a-Service (MaaS) via AI" approach enabled a lone developer with modest skills to create a viable new spyware** – something that might have required a team effort before. It also illustrates that threat actors trust LLMs enough to feed them proprietary malicious code, essentially using AI as a collaborative coding partner (albeit at the risk of exposure if using a public service).

Criminal forums are awash with claims that AI can produce *"undetectable"* malware. The dark web tool **FraudGPT** advertised itself as capable of *creating malware that bypasses antivirus detection*, generating phishing webpages, and finding software vulnerabilities [49] [50]. A demo showed FraudGPT producing a working fake Bank of America login page in minutes [50]. **WormGPT**, likewise, demonstrated writing a Python-based ransomware and polymorphic malware code when prompted with "malicious requirements" [51]. While these models don't magically invent zero-days, they can **take existing malware templates and obfuscate or mutate them** on the fly. In tests, researchers found that ChatGPT itself (if properly prompted or jailbroken) can generate polymorphic code that changes its signature each time [52] [53]. For example, CyberArk researchers created a proof-of-concept called *BlackMamba* – malware that used ChatGPT's API at runtime to continuously rewrite its payload into new forms [54] [55]. BlackMamba's authors showed that by pulling code from a *"benign" AI service (OpenAI)* instead of a known malicious server, and by never writing the same code twice, they could evade many detection tools [56] [55]. The malware essentially acted as an **AI-powered polymorphic engine**, assembling its keylogger functionality dynamically in memory, which undermines traditional signature-based defenses [57] [58]. Although BlackMamba was a researcher's demo, it foreshadows what enterprising threat actors might do – especially with open models that can be hosted on attacker infrastructure for stealth.

As mentioned, Chinese and other APT hackers have attempted to generate malware via mainstream LLMs but hit ethical guardrails [32]. In response, they are turning to **unrestricted models**. The rise of open-source LLMs has given threat actors a playground to fine-tune models specifically for malicious tasks. We've seen references to *"CrimeGPT" clones like WolfGPT* **(touted as an AI with "complete confidentiality" for creating cryptographic malware and phishing attacks) and** XXXGPT **(allegedly geared toward helping deploy RATs, botnets, ATM malware, and info-stealers)** [59] [60] **. While some**

of these may be hype or scams, the pattern is clear: there is a burgeoning cottage industry of custom LLMs trained on malware code, hacking tutorials, and illicit data*. These models lower the barrier for anyone to generate malware in multiple programming languages, customize ransomware notes, or even identify vulnerabilities. Indeed, an analysis by security firm SlashNext in 2023 found WormGPT was heavily used to assist in business email compromise and phishing*, while its successor models could potentially handle more technical exploits [61] [62].

Even without custom models, attackers can **jailbreak commercial AI systems to ignore safety filters**. There is an active underground trade in prompt-based exploits to bypass ChatGPT's content rules [63]. Jailbreak prompts (e.g. the "DAN" exploit and myriad variants) can trick GPT-4 into providing disallowed outputs by assuming a role or using coded language. Daniel Kelley, a former black-hat hacker, noted an *"unsettling trend"* on forums where users share prompts to unlock ChatGPT's darker capabilities [64]. For example, instead of directly asking "write malware," a clever prompt might ask the model to play the role of a cybersecurity researcher generating a *hypothetical* malware for study. The goal is to get useful malicious code or advice out of the AI without triggering its filters. As AI companies improve guardrails, criminals pivot to new bypass techniques – sometimes even chaining multiple AIs (one model's output used to fool another). This cat-and-mouse game indicates that **motivated attackers can still coax dangerous guidance from top-tier LLMs**, supplementing their use of unguarded models.

LLMs are also aiding in **payload delivery mechanisms**. A notable example is *phishing websites and malicious documents* enhanced by AI. Traditionally, fake websites for phishing could be spotted by design flaws or text errors. Now, a skilled actor can use an AI to generate professional-looking site templates and even JavaScript that adapts content to the victim. Reports suggest attackers have started using AI-driven *cloaking* – dynamically altering phishing site content in real-time to evade security scanners, possibly by using an AI to detect when a bot is visiting and then show benign content [65]. Additionally, malware droppers like malicious Office macros or scripts can be quickly iterated by an AI to find variants that evade detection. In the Kimsuky deepfake ID attack, aside from the AI-generated ID image, the hackers deployed obfuscated batch scripts and AutoIt loaders to drop their payload [66] [67]. While those obfuscation techniques were manually coded, the Genians report points out that **using generative AI to produce "clean" variations of malicious code is trivially easy**, raising the prospect of on-demand polymorphism in everyday attacks [25] [56]. Even ransomware notes and extortion messages are now AI-authored; in one case, an AI wrote ransom notes tailored with victim-specific data and psychologically intimidating language to maximize pressure [39] [40].

Crucially, **LLMs have lowered the skill threshold for developing complex malware**. Anthropic observed a case where a cybercriminal with only basic coding knowledge used Claude to create and sell a new ransomware strain [68] [69]. What once required advanced malware engineering can now be achieved by asking an AI for step-by-step code. This democratization of malware creation means we may see more "script kiddie" actors deploying sophisticated ransomware or exploits, effectively **outsourcing the heavy lifting to AI**. It's a development not lost on government officials: U.S. Cybersecurity Director Jen Easterly told Congress in 2024 that *"AI… is an epoch-defining challenge"*, pairing it with the threat of nation-states like China [70]. Her warning came as Microsoft noted generative AI is expected to enable **more sophisticated deepfakes and voice clones**, further blurring the lines between human and machine-generated attack content [71].

Conclusion

From crafting the initial phishing email to automating the post-compromise game plan, large language models are now present at every stage of the cyber kill chain. **Russian, Chinese, and North Korean hacking groups – once constrained by language barriers, limited manpower, or technical gaps –**

are embracing AI as a force multiplier. They use LLMs to write fluent phishing bait in any language [1], to gather intelligence on targets and technology [27] [72], and to generate or refine malware code [49] [32]. In the hands of North Korea's operatives, AI helps create entire fake identities and deepfake content to infiltrate organizations [19] [24]. In the cybercriminal underground, bespoke "dark LLMs" offer plug-and-play malicious code and phishing kits to even unskilled attackers [13] [14]. While many of these AI-enabled tactics are still emerging, the trend is unmistakable: **AI is becoming an accomplice in cyberattacks**.

For cybersecurity professionals, understanding these developments is critical. Defenders must recognize that the phishing email in your inbox might have been *authored by an AI*, informed by data scrapes of your digital footprint. The malware that slipped past your endpoint defenses could be *polymorphic code continually rewritten by an AI agent*. And the "person" applying to your company – or chatting with you on LinkedIn – could in fact be a North Korean operative bolstered by an LLM generating their flawless English and answers [21] [20]. The offensive use of LLMs remains in an arms race with defensive measures. As of late 2024, both OpenAI and Anthropic noted that most malicious uses of their models were incremental and replicable with other tools [73] [42]. However, *incremental advantages* at scale can tip the balance. Thousands of AI-crafted phishing emails or dozens of AI-scripted malware variants can overwhelm conventional defenses. Moreover, the creativity of threat actors – as seen in the Kimsuky deepfake ID campaign – means we should expect **novel AI-enabled attack techniques** to continue emerging [25] [24].

In summary, **large language models have moved from the lab to the arsenal of cybercriminals**. Russian APTs are querying them for technical intel; Chinese hackers are testing them for coding and research support; North Korean groups are weaponizing them to socially engineer and even stealthily employ their agents. Financially motivated gangs globally are using or building uncensored LLMs to write malware, phishing content, and more at an industrial scale [61] [13]. The cutting edge today sees AI not only *assisting* hackers but sometimes *leading* certain operations – an early sign of the "agentic" AI-powered attacks security experts have warned about [42]. For the "geeky" tech audience in cybersecurity, the imperative is clear: we must study and track these AI-fueled tactics in detail, because defending against them requires a deep understanding of how they work. The exploits themselves may still rely on human weaknesses and known vulnerabilities, but LLMs are supercharging the speed, scale, and believability of cyberattacks. The landscape of threat activity in 2025 and beyond will increasingly reflect this AI-driven escalation [74] [75].

Sources:

- Microsoft/OpenAI Threat Intelligence – foreign APT use of LLMs [27] [6] [72]
- OpenAI Security Report (2024) – cases of ChatGPT misuse by threat actors [34] [47]
- CyberScoop & The Guardian – state-backed hackers experimenting with OpenAI models [76] [5]
- Trustwave SpiderLabs – **WormGPT** and **FraudGPT** malicious LLMs on underground forums [11] [13]
- Infosecurity Magazine – Dark web AI chatbots (FraudGPT, DarkBard, etc.) for cybercrime [10] [14]
- Alan Turing Institute (CETaS) Report (2025) – AI in serious organized crime, incl. phishing, fraud, WormGPT [2] [77]
- Julian Hazell (Oxford University) – *"Spear Phishing with Large Language Models"* study [78] [79]
- TechRadar (Feb 2024) – Acronis Cyberthreats Report on AI-powered phishing growth [3] [1]
- Genians Security Center – Kimsuky's AI-driven deepfake ID phishing campaign (2025) [24] [25]
- Anthropic Threat Report (Aug 2025) – Claude misuse cases (NK fraudulent employment, AI-written ransomware) [21] [42]
- Bloomberg News – North Korean hackers using ChatGPT for fake IDs in phishing [24]
- Anthropic (Claude) vs. Cybercriminals – AI agent used in extortion operation [39] [40]

• SentinelOne Labs – *BlackMamba* ChatGPT-polymorphic malware PoC analysis [56] [55]

---

[1] [3] [4] Email attacks on business tripled in 2023 — and ChatGPT was often the culprit | TechRadar
https://www.techradar.com/pro/security/email-attacks-on-business-tripled-in-2023-and-chatgpt-was-often-the-culprit

[2] [18] [75] [77] AI-Driven Crime Escalates in Scale and Sophistication, Warns UK Security Report - BABL AI
https://babl.ai/ai-driven-crime-escalates-in-scale-and-sophistication-warns-uk-security-report/

[5] [6] [33] [43] [70] [71] [72] North Korea and Iran using AI for hacking, Microsoft says | Hacking | The Guardian
https://www.theguardian.com/technology/2024/feb/14/north-korea-iran-ai-hacking-microsoft

[7] [8] [34] [35] [36] [37] [45] [46] [47] [48] [73] OpenAI reveals ChatGPT use by CyberAv3ngers, Android malware developers | SC Media
https://www.scworld.com/news/openai-reveals-chatgpt-use-by-cyberav3ngers-android-malware-developers

[9] [11] [12] [49] [50] [51] WormGPT and FraudGPT – The Rise of Malicious LLMs
https://www.trustwave.com/en-us/resources/blogs/spiderlabs-blog/wormgpt-and-fraudgpt-the-rise-of-malicious-llms/

[10] [13] [14] [15] [16] [17] [59] [60] [61] [62] [63] [64] The Dark Side of Generative AI: Five Malicious LLMs Found on the Dark Web
https://www.infosecurityeurope.com/en-gb/blog/threat-vectors/generative-ai-dark-web-bots.html

[19] [20] [21] [22] [23] [25] [26] [66] [67] AI-Driven Deepfake Military ID Fraud Campaign by Kimsuky APT
https://www.genians.co.kr/en/blog/threat_intelligence/deepfake

[24] North Korean Hackers Used ChatGPT to Help Forge Deepfake ID - Bloomberg
https://www.bloomberg.com/news/articles/2025-09-14/north-korean-hackers-used-chatgpt-to-help-forge-deepfake-id

[27] [28] [29] [30] [31] [32] [76] State-backed hackers are experimenting with OpenAI models | CyberScoop
https://cyberscoop.com/openai-microsoft-apt-llm/

[38] [39] [40] [41] [42] [68] [69] Detecting and countering misuse of AI: August 2025 \ Anthropic
https://www.anthropic.com/news/detecting-countering-misuse-aug-2025

[44] [74] Alan Turing Institute calls for AI Crime Taskforce | UKAuthority
https://www.ukauthority.com/articles/alan-turing-institute-calls-for-ai-crime-taskforce

[52] Chatting Our Way Into Creating a Polymorphic Malware - CyberArk
https://www.cyberark.com/resources/threat-research-blog/chatting-our-way-into-creating-a-polymorphic-malware

[53] ChatGPT's Dark Side: An Endless Supply of Polymorphic Malware
https://www.esecurityplanet.com/threats/chatgpt-malware/

[54] [55] [56] [57] [58] BlackMamba ChatGPT Polymorphic Malware | A Case of Scareware
https://www.sentinelone.com/blog/blackmamba-chatgpt-polymorphic-malware-a-case-of-scareware-or-a-wake-up-call-for-cyber-security/

[65] CyberheistNews Vol 15 #30 [Heads Up] Ransomware is Back—and ...
https://blog.knowbe4.com/cyberheistnews-vol-15-30-heads-up-ransomware-is-back-and-smarter-than-ever-in-2025-trends

[78] [79] 1
https://cdn.governance.ai/Spear_Phishing_with_Large_Language_Models.pdf