Artificial Intelligence and the New Economics of Cyberattacks

Vaibhav Garg, Jayati Dev[1]

Why is information security hard? This question is in part answered by the underlying Economics of attack and defense. Different economic models result in distinct equilibria, some favor attackers, whereas others advantage defenders. These equilibrium positions may change with the introduction of powerful new Large Language Models (LLMs). In this paper we investigate three key impacts of these AI-based disruptions. First, emerging Artificial Intelligence (AI) capabilities may increase both the number of attackers as well as the types of cybercrime they engage in. Second, we discuss how these capabilities will be developed by a small set of players, which will potentially increase concentration in cybercriminal markets. Third, finally as both attackers and defenders attempt to use AI, the asymmetry between attackers and defenders will decrease. Thus, we show how AI will impact the economic functioning of cybercrime markets.

[1] Cybersecurity Research and Development, Comcast.

1. Introduction

In Isaac Asimov's famous Foundation Series [1], the protagonist Harry Sheldon was able to model, and thus predict, humanity's future until the emergence of a new unforeseen 'technology' in the form of mutants. Similarly, the emergence of Generative AI and associated Large Language Models (LLMs) threaten to disrupt our traditional understanding of Cybersecurity Economics, particularly the models of cyberattack and cyber defense. AI may therefore fundamentally reshape the structure of cybercriminal organizations and markets. This transformation may then in turn impinge or assist defender efforts.

Consider that LLMs may be used to generate more convincing social engineering attacks (e.g., phishing emails), reduce the effort to build malware (e.g., by automatically writing code), or be used to generate attacks by highly sophisticated persistent threat actors, previously exclusive in the domain of nation state attackers according to recent research by Nickson Quak from the Cyber Threat Alliance [2]. According to Lin et al., attackers may jailbreak mainstream LLMs or alternatively they may leverage malicious LLMs designed for nefarious purposes, e.g., WormGPT, FraudGPT [3]. These developments create the impression that AI fundamentally advantages the attacker. Yet AI has traditionally been the tool of cyber defenders. AI-based tools have been used in cybersecurity for spam/phishing detection, intrusion detection, malware detection, vulnerability scanning and remediation, etc. [4]. The market for AI in cybersecurity, according to Taddeo et al., is expected to grow by an order of magnitude by 2025 compared to 2016 [5]. Prior work by Gupta et al. highlights how emerging AI like LLMs also promise to advance defender capabilities [6]. LLMs can be used to reduce the amount of technical expertise needed by a Security Operations Center (SOC) analyst by collating multiple indicators of compromise and providing the combined intelligence in a human readable format. They can be used to conduct code reviews for security. They can also be used to generate patches more quickly for known vulnerable code.

What then is the overall impact of AI on the dynamics between attackers and defenders? In this position paper we attempt to address this by looking at three questions. First, what is the impact of AI on organized cybercrime? Second, what is the impact of AI on reliability of cybercrime markets? Third, what is the impact of AI on attacker and defender capabilities? We make three corresponding contributions investigating the overall impact of AI on the dynamics between attackers and defenders:

1. Exemplify how AI may increase both the number of cybercriminals and the types of cybercrimes they engage in.
2. Demonstrate how economic forces of AI-powered tools may cause further concentration of power in the cybercriminal market within an even smaller set of actors (i.e., cybercriminal organizations).
3. Present a case of how AI may decrease the asymmetry between attackers and defenders regardless of the model.

The rest of the paper is organized as follows. Section 2 discusses organization in cybercrime. In Section 3, we reframe the question of system reliability for cybercrime products through relevant economic models. Section 4 presents the two canonical models of attack and defense. In section 5 we discuss how AI may inform the economic models discussed in Sections 3 and 4. Section 6 concludes with a discussion of future work.

2. Rise of Organized Cybercrime

Why do individuals engage in cybercrime? One explanation is that the entry barrier for engaging in criminal activity is lower than that for engaging in legitimate activity [7, 8]. However, a recent report from the National Security Telecommunications Advisory Committee (NSTAC) claims that investments in cybersecurity by defenders have made it more difficult to engage in cybercrime [9]. Consequently, while the marginal cost of cybercrime has gone up the marginal benefits have gone down. It can then be argued that cybercriminals have responded in line with Coase's theory of the firm to become organized and thus also become specialized [10].

Prior research by Garg et al. notes that cybercriminals typically organize as either gangs or mobs [11]. Gangs are characterized by a central leader. The organization size of gangs is impinged by Dunbar limits [12], which is typically 100-230 members. Such gangs specialize in a narrow set of cybercrimes to leverage competitive advantage. In contrast, mobs are characterized by a complex hierarchy with multiple leaders so that there is no single critical link that can collapse. Organization size of mobs have a larger portfolio of cybercriminal activity.

Additional organization and cooperation for these organized cybercrime enterprises are driven by the technical complexity of conducting cybercrime. Consider that to run a successful ransomware campaign, an attacker will need access to phishing infrastructure, access to malware authors or pre-packaged ransomware along with decryption keys, and a support service agent who can respond to queries from infected "customers". Arguably, as Collier et al. mention, running a cybercriminal enterprise is akin to running a standard IT company [13].

3. Cybercrime's System Reliability Problem

Cybercriminal organizations can thus be treated as individual 'companies' that produce competing or complementary products. In the former case, they may integrate horizontally to increase market size and benefit from economies of scale. In the latter case they may integrate vertically to benefit from process efficiencies. These integrations may allow cybercriminal s*eller o*rganizations in a cybercrime marketplace to differentiate their products more easily to cybercriminal b*uyer o*rganizations [14].

The customer focus on these markets means that *sellers* need to ensure that the products offered as Cybercrime-as-a-Service (CaaS) on sale are reliable [15]. Considering these CaaS offerings as a system, we can apply Varian's [16] three distinct models for system reliability.

Consider a market where various stages of a cybercrime product's supply chain are controlled by distinct agents. For example, for a ransomware product one agent may provide the phishing infrastructure, another agent may provide the malware, and a third agent may provide the customer service. The success of the ransomware product in this case would be controlled by the least reliable element of the supply chain or the weakest link. In this scenario the agent with the least benefit cost ratio controls the system reliability. More importantly, as Varian [16] notes, increasing the number of agents decreases the reliability of the system. Thus, to produce a more reliable cybercriminal product – cybercriminal organizations will aim to consolidate the various stages of cybercrime tools production and absorb functions of two or more of its supply chain agents. This would lead to greater vertical integration.

Consider a different market where products are vertically integrated; however, multiple cybercriminal organizations offer competing products as Anderson et al. points out [17]. In this case the reliability of the cybercriminal product will be determined by the best effort model, i.e., the product ecosystem is as reliable as the most reliable product on the cybercriminal market. Unsurprisingly in this case reliability will be decided by the agent with the highest benefit-cost ratio. Consequently, agents with a lower benefit-cost ratio will produce a less reliable product or would not produce products at all. If they do produce a less reliable product, due to its lower reliability, it will be ignored by the customers in lieu of more reliable ones. This condition occurs assuming that there is no information asymmetry between the seller and the buyer as well as equivalent pricing between the two products. Hence, agents with a lower benefit-cost ratio will either leave the market or be absorbed by more successful agents (in the best effort model) so that the latter has more buyers. The cybercrime market will then gravitate towards horizontal integration.

Varian highlights a third model where individual cybercriminals may cooperate to peer produce cybercriminal products. Like Langlois and Garzarelli's discussion of open-source production of tools for legitimate use cases [18], system reliability in this case depends on the total effort of all the cybercriminals, but it will be decided by the agent with the highest benefit-cost ratio (the best effort case).

4. Attack versus Defense

In the previous section, we discussed how economic factors drive the evolution of organized cybercrime and associated markets. Eventually these cybercriminals and their tools must face defenders. The success of cybercriminals in this case is combination of the economics of attack and defense. The consensus, as highlighted by Ross Anderson's work [19] before

LLMs were commonly available for cybercrime, was that the economics of cybercrime favors the attacker.

Consider the ransomware example introduced in the earlier section. Let us assume a model in which an attacker sends a phishing email to all the employees at a company. Any employee who clicks on the link in the phishing email can cause the ransomware to execute and cause a security incident at the company. The attacker in this case needs to only find one employee who makes an error. The defender needs to make sure that all employees have a 100% rate of finding and mitigating phishing emails. The attacker can concentrate its resources, while the defender must spread them out. Thus, the advantage under this model is to the attacker.

However, this model would also depend on an organization's maturity. Let us consider a different model in which an organization implements Zero Trust Architecture [28], i.e., where individuals are not trusted by default and verification is needed before anyone can access the organization's resources. For such an organization, errors on one employee's end would need to be supplemented with several other cyberattacks working in tandem to ensure the whole organization falls victim. Indeed, most successful breaches require a sequence of attacks to execute one after the other in what is often described as a Cyber Kill Chain [20, 21]. Thus, the attacker needs to get lucky multiple times in a row, while the defender needs to get lucky once in the chain as Slayton points out [22]. This is the economic argument to support the implementation of Zero Trust.

5. AI Impacts

In the previous sections we discussed how Economics drives organization in cybercrime, the reliability of cybercrime products, and the relative advantage between attackers and defenders. While the insights from these models have been useful, they will need to be reconsidered now to address the impact of AI.

It is clear from the rise of FraudGPT and WormGPT, alongside legitimate but jailbroken LLMs, that AI will make it easier to engage in cybercrime. Chui et al. have argued that AI will lead to greater productivity [23]; surely, this may apply to cybercriminal productivity as well [24]. First, AI may reduce the amount of technical sophistication that is required to conduct cybercrime. For example, according to a recent survey paper, an attacker may use LLMs to help them generate malware [25]. Second, it may improve the effectiveness of certain cyberattacks. For example, a recent report says that an attacker may use LLMs to generate more targeted, and therefore more compelling, phishing lures [26]. The former will reduce the cost of doing cybercrime, while the latter will increase the benefits, i.e., AI may increase overall profits from cybercrime. Thus, more individuals may be incentivized to engage in cybercrime. Furthermore, as AI reduces the technical expertise to engage in cybercrime, existing gangs and mobs may expand the types of cybercrime they engage in.

AI may even impact the evolution of cybercriminal markets. As noted in Section 3, longer supply chains or increasing number of agents in a cybercriminal product supply chain may reduce reliability. The current technological landscape has created the need for multiple agents, as investments in cyber defense have made it harder for cyberattacks to succeed. Cybercriminals must thus plan sophisticated campaigns, each stage of which requires a specialized cybercriminal skill set. However, LLMs may reduce the amount of expertise required going forward. Just as the advent of computers reduced the demand for typists, LLMs will allow cybercriminal organizations to bring certain parts of the operation in-house. For example, organized cybercrime – much like legitimate companies – may use LLMs to provide customer support. Thus, AI may help drive vertical integration in cybercriminal markets.

With more cybercriminal organizations engaging in an ever-increasing number of cybercriminal activities, with greater revenue enabled by greater system reliability and other efficiencies enabled by vertical integration, we may witness an increase in competition between these groups to grow their individual market shares. As Varian notes [16], under this dynamic system reliability depends on the agent with the highest benefit cost ratio while other agents freeride. One way this outcome can materialize may be in the form of a small set of cybercriminal organizations in the form of mobs. These mobs would provide a comprehensive set of LLM-enabled cybercrime products, both capabilities and infrastructure. The remaining vast majority of cybercriminal organizations stay at the level of gangs who are just customers for the mob.

As these mobs and gangs use their AI and LLM enabled capabilities to conduct cyberattacks, defenders will look to leverage the same technologies to scale defense. As noted in Section 4, the classical model of attack versus defense favors the attacker; the defender must spread their resources to defend everywhere – the attacker need only find one undefended resource. AI may help scale both defender and attacker capabilities. However, as the attacker was already in an advantageous position any marginal gains for attack should be less than the marginal gains for the defender.

The reverse may be true for a Zero Trust based attack versus defense model. Here the defender has the advantage, as the attacker needs to get lucky multiple times, and the defender needs to be lucky only once. In this case too AI may further both attacker as well as defender capabilities. However, the marginal gains for the more favored party, i.e., attacker, should be greater than marginal gains for the less favored party, i.e., defender. Overall, we hypothesize that the introduction of AI to attack and defense would reduce the asymmetry between the two regardless of the underlying model.

6. Conclusion and Future Work

When it comes to cybersecurity, is AI good or bad? This is the question that everyone, regulators, industry professionals, academics, and civil society hopes to answer. This is the

puzzle that the White House Executive Order on AI aims to solve when it asks the Department of Homeland Security to conduct a cross-sector risk assessment on AI for critical infrastructure [27]. This is the worry that underlies viral coverage of FraudGPT and WormGPT. This is the conundrum that keeps cyber executives open to receiving calls from vendors promising the newest AI cure-all.

In this position paper we examine the impact of AI on cybercrime on three dimensions: (1) cybercrime organization, (2) reliability of cybercrime markets, and 3) asymmetry between attackers and defenders. Our initial examination suggests that AI may reduce the barrier to engage in cybercrime; thus, we have more cybercriminals engaging in more kinds of cybercrime. Furthermore, AI may increase concentration in cybercrime markets through both vertical and horizontal integration. Finally, AI may reduce the asymmetry between attackers and defenders due to the broad range of tools available to both groups.

In future, we intend to build on this initial observation through experimentation. Furthermore, while there is an increasing interest in AI tools available to attackers, the utility of such tools used by defenders is not measured as much. We would like to explore the impact of AI on organization in defense and associated markets. Additionally, there is limited modeling done on establishing how attackers are at an advantage in cybercrime ecosystems. We also hope to increase the maturity of our analyses by including formal mathematical models. Finally, we intend to include implications for technologists and policy makers as AI changes the landscape of threats.

References

[1] Asimov, I., 2010. Foundation, Foundation and Empire, Second Foundation: Introduction by Michael Dirda. Everyman's Library.

[2] Quak, N., 2023. How Emerging Technologies Threaten Our Cybersecurity. Cyber Threat Alliance, White Paper.

[3] Lin, Z., Cui, J., Liao, X. and Wang, X., 2024. Malla: Demystifying Real-world Large Language Model Integrated Malicious Services. arXiv preprint arXiv:2401.03315.

[4] Sarker, I.H., Furhad, M.H. and Nowrozy, R., 2021. AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. *SN Computer Science*, *2*(3), p.173.

[5] Taddeo, M., McCutcheon, T. and Floridi, L., 2019. Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword. *Nature Machine Intelligence*, *1*(12), pp.557-560.

[6] Gupta, M., Akiri, C., Aryal, K., Parker, E. and Praharaj, L., 2023. From ChatGPT to ThreatGPT: Impact of Generative AI In Cybersecurity and Privacy. *IEEE Access*.

[7] Garg, V. and Camp, L.J., 2015. Why Cybercrime? *ACM SIGCAS Computers and Society*, *45*(2), pp.20-28.

[8] Garg, V., Husted, N. and Camp, J. 2011. The Smuggling Theory Approach to Organized Digital Crime. In *2011 eCrime Researchers Summit* (pp. 1-7). IEEE.

[9] 2024. Measuring and Incentivizing the Adoption of Cybersecurity Best Practices. *National Security Telecommunications Advisory Committee (NSTAC)*.

[10] Coase, R.H., 1937. The Nature of the Firm. *Economica*, *4*(16), pp.386-405.
[11] Garg, V., Afroz, S., Overdorf, R. and Greenstadt, R., 2015. Computer-Supported Cooperative Crime. In *Financial Cryptography and Data Security: 19th International Conference, FC 2015, San Juan, Puerto Rico, January 26-30, 2015, Revised Selected Papers 19* (pp. 32-43). Springer Berlin Heidelberg.
[12] Brenner, S.W. 2002. Organized Cybercrime-How Cyberspace May Affect the Structure of Criminal Relationships. North Carolina Journal of Law & Technology 4, 1 (2002)
[13] Collier, B., Clayton, R., Hutchings, A. and Thomas, D., 2021. Cybercrime Is (Often) Boring: Infrastructure and Alienation in a Deviant Subculture. The British Journal of Criminology, 61(5), pp.1407-1423.
[14] Van Wegberg, R., Miedema, F., Akyazi, U., Noroozian, A., Klievink, B. and van Eeten, M. 2020. Go See a Specialist? Predicting Cybercrime Sales on Online Anonymous Markets from Vendor and Product Characteristics. In Proceedings of the Web Conference 2020 (pp. 816-826).
[15] Akyazi, U., van Eeten, M.J.G. and Ganan, C.H., 2021. Measuring Cybercrime as a Service (CaaS) Offerings in A Cybercrime Forum. In Workshop on the Economics of Information Security.
[16] Varian, H., 2004. System Reliability and Free Riding. In Economics of information security (pp. 1-15). Boston, MA: Springer US.
[17] Anderson, R., Clayton, R., Böhme, R. and Collier, B. 2021. Silicon Den: Cybercrime is Entrepreneurship. In *Workshop on the Economics of Information Security*.
[18] Langlois, R.N. and Garzarelli, G., 2014. Of Hackers and Hairdressers: Modularity and The Organizational Economics of Open-Source Collaboration. In *Online Communities and Open Innovation* (pp. 11-29). Routledge.
[19] Anderson, R., 2001, December. Why Information Security is Hard - An Economic Perspective. In Seventeenth Annual Computer Security Applications Conference (pp. 358-365). IEEE.
[20] 2011. The Cyber Kill Chain. Lockheed Martin. https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html
[21] Dargahi, T., Dehghantanha, A., Bahrami, P.N., Conti, M., Bianchi, G. and Benedetto, L., 2019. A Cyber-Kill-Chain based Taxonomy of Crypto-Ransomware Features. Journal of Computer Virology and Hacking Techniques, 15, pp.277-305.
[22] Slayton, R., 2016. What is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment. International Security, 41(3), pp.72-109.
[23] Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L. and Zemmel, R., 2023. The Economic Potential of Generative AI: The Next Productivity Frontier. McKinsey.
[24] Torre, Richard De La. 2023. How AI is Shaping the Future of Cybercrime. Dark Reading. https://www.darkreading.com/vulnerabilities-threats/how-ai-shaping-future-cybercrime.
[25] Kaloudi, N. and Li, J., 2020. The AI-based Cyber Threat Landscape: A Survey. ACM Computing Surveys (CSUR), 53(1), pp.1-34.
[26] Violino, B. 2023. AI Tools Such as ChatGPT Are Generating a Mammoth Increase in Malicious Phishing Emails. CNBC. https://www.cnbc.com/2023/11/28/ai-like-chatgpt-is-creating-huge-increase-in-malicious-phishing-email.html.

[27] 2023. Executive Order on The Safe, Secure, And Trustworthy Development and Use of Artificial Intelligence. The White House. The United States Government. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/
[28] Stafford, V.A., 2020. Zero Trust Architecture. NIST Special Publication, 800, p.207.