VIEWPOINT

ARTHUR D LITTLE
2024

AI IN CYBERSECURITY

How companies can navigate cyber threats & exploit opportunities in the age of AI

AI is rapidly evolving from a set of complex expert tools to a user-friendly technology that's substantially impacting the cybersecurity landscape. Even as AI enhances threat detection and response, it provides cybercriminals with multiple new attack methods. Additionally, when AI systems within organizations are not properly secured, they can introduce new vulnerabilities. This Viewpoint explores the dual nature of AI in cybersecurity and helps companies understand how to defend themselves against emerging threats.

AUTHORS

Maximilian Scherr

Igor Stepanov

Daria Bryzhytska

Pawel Jablonski

Michael Papadopoulos

Tom Teixeira

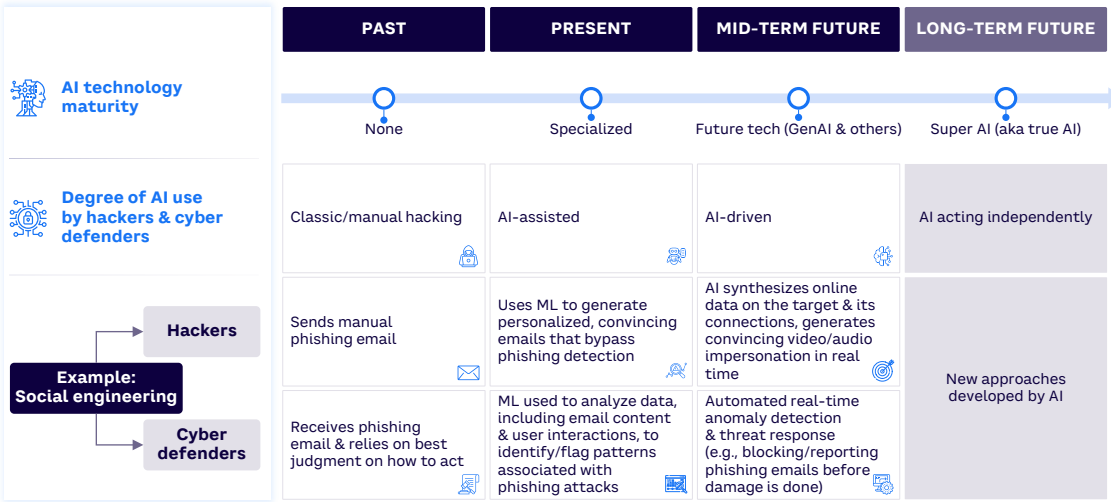
AI IS TRANSFORMING CYBERSECURITY

Imagine you're the COO of a production company and you get a call from the head of procurement asking you to approve an urgent purchase to ensure your business continues to run smoothly. Because you understand the urgency and have worked with the person for many years, you do not question the request. A few days later, you meet with her and ask whether everything worked out with the purchase — but she has no idea what you're talking about. Suddenly, you realize the call quality wasn't that good and came through a third-party messenger app, although it was definitely her voice and usual speech patterns. You couldn't see her face, but you didn't need to because there was nothing suspicious about the call at that moment. Of course, it is now too late: the transaction went through, and it's irreversible. This is not fiction — it happened to a finance worker at a multinational company earlier this year.

AI IS ALREADY AN INTEGRAL PART OF MODERN TECHNOLOGY

AI is at a turning point, shifting from specialized machine learning (ML) and other techniques that require skills training to versatile, user-friendly techniques based on generative AI (GenAI). AI is already an integral part of modern technology, and cybersecurity is no exception. In fact, AI can be leveraged for both defensive and offensive actions: improving organizations' ability to detect and respond to cyber threats while giving cybercriminals tools to launch more sophisticated and targeted attacks. AI-based tools introduced into corporate IT ecosystems can also be a source of new vulnerabilities if they're not integrated using stringent security measures. This increases the organization's attack surface, making it more vulnerable to cyber threats (see Figure 1).

Figure 1. AI-based hacking shifting from highly specialized machine learning to future AI technologies



Source: Arthur D. Little

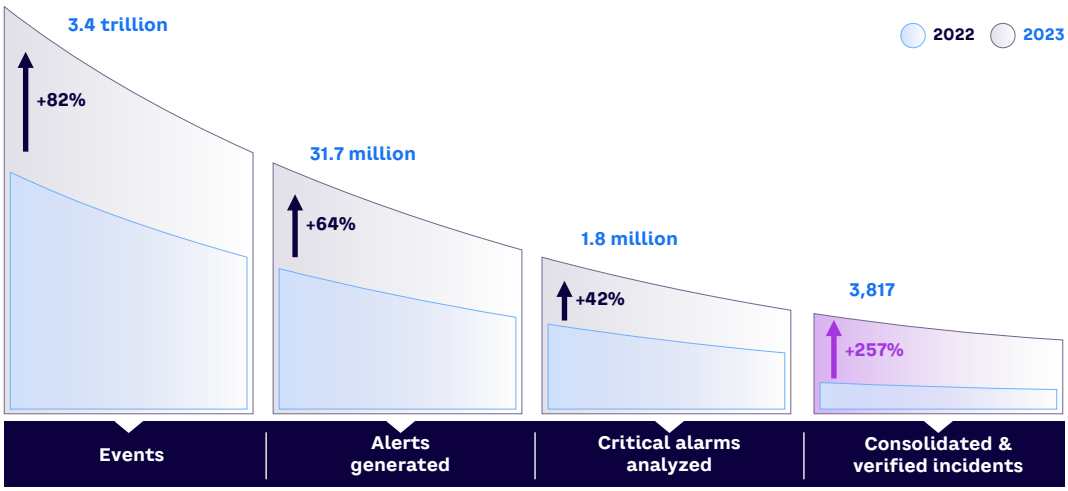
HACKERS ARE WEAPONIZING AI

Current AI developments give hackers a strong advantage, helping them launch more cyberattacks that pose significant threats to individuals, businesses, and governments worldwide. Security professionals are seeing the impact: CANCOM reports a 257% surge in consolidated and verified incidents in 2023 (see Figure 2). During the same period, events grew by 82% (alerts and critical alarms grew by 64% and 42%, respectively). This shows that despite cyber-defense mechanisms, on average, hackers have been more successful (the number of confirmed incidents grew significantly faster than the underlying events). Furthermore, according to a Deep Instinct survey, 75% of security professionals in the US witnessed an increase in attacks over the past 12 months, with 85% of those professionals claiming the rise comes from bad actors using GenAI.

Arthur D. Little (ADL) has identified some common patterns in AI integration, including how it can change a company's business model (see Figure 3). In the first wave, AI can increase efficiency through automation. Next, it increases effectiveness by enhancing skills. Finally, it's a catalyst for creating new business models. We see the same pattern in the development of AI-enabled security threats:

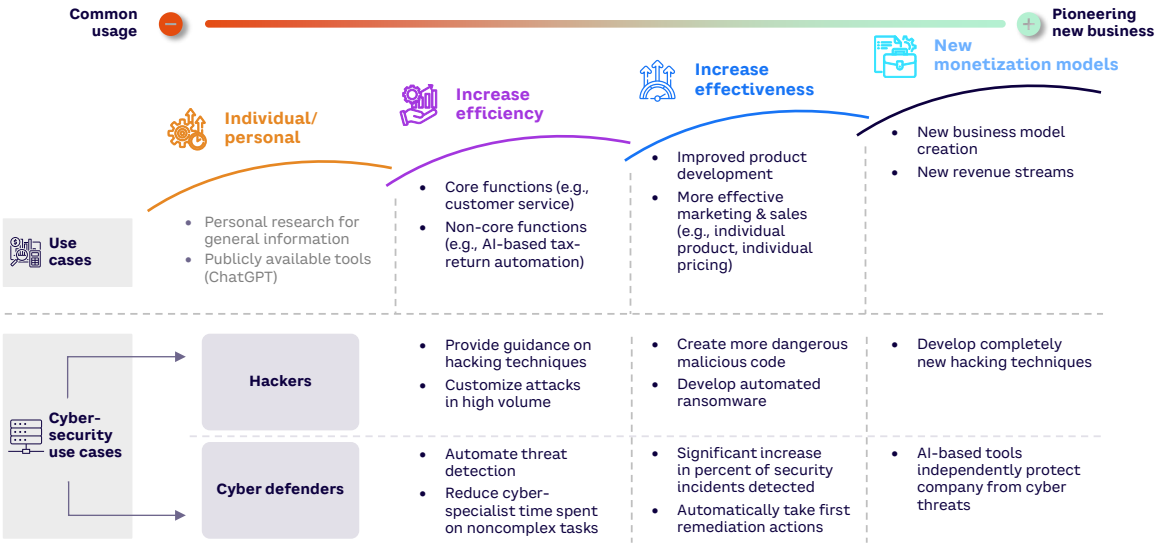
1. **Efficiency.** AI techniques have already made hackers more efficient through automation and optimization. Recently released AI-based tools let hackers quickly generate malicious code and help them conduct tailored hacking attacks on specific environments.
2. **Effectiveness.** AI-enabled attacks are increasing effectiveness through enhanced social engineering techniques, AI-coordinated lateral movements in hacked environments, and adaptive malware. These attacks are also more sophisticated; for example, social engineering attacks are more personalized to their target and can adapt in real time (making detection more difficult).
3. **Transformation.** AI will transform cyberattacks by creating completely new capabilities and adaptability levels. As with generative adversarial network (GAN) malware, AI-based malware can rapidly change to avoid detection by automated tools. Worse, it has high levels of autonomy and self-development; more advanced versions might only require a knowledge of the hacker's goal — once delivered to the system, it would find ways to achieve the goal on its own.

Figure 2. Surge in incidents, 2022–2023



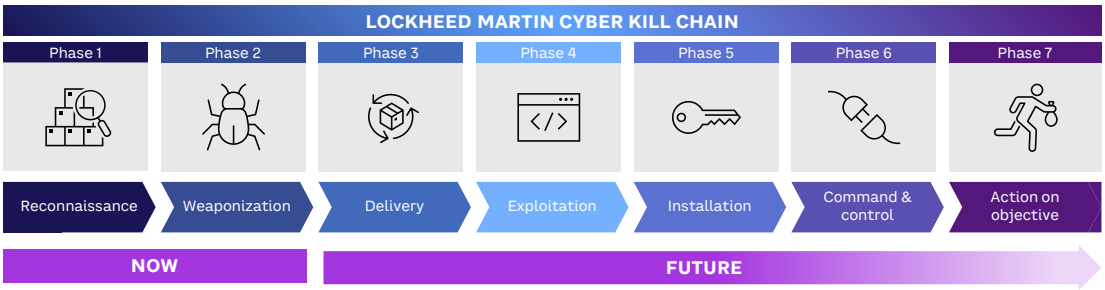
Source: Arthur D. Little, CANCOM

Figure 3. AI integration patterns



Source: Arthur D. Little

Figure 4. Stages of a cyberattack



Source: Arthur D. Little

Currently, AI is mainly being used to make hackers more efficient in the early stages of an attack. In the industry-standard Lockheed Martin Cyber Kill Chain shown in Figure 4, we see that AI-based tools are being used in the reconnaissance stage and (to some extent) in the weaponization stage. We expect to see more diverse and smarter tools in the near future; DeepLocker is already launching tools that can be used in the delivery stage.

The most common AI-driven attack type is social engineering, comprising 83% of all cyberattacks. The danger level has increased with AI: Darktrace research reported a 135% increase in "novel social engineering attacks" among thousands of active email-protection customers from January to February 2023 (ChatGPT was released on 30 November 2022). This threat is expected to increase: Google's "Cloud Cybersecurity Forecast 2024" report predicts phishing will be significantly enhanced by AI in 2024.

New, widely available deepfake technologies are emerging that create a close replication of a person's voice based on a small audio sample, making impersonation extremely accessible. Deepfake technology, especially voice emulation, can create highly personalized, targeted attacks based on victim profiles, optimizing timing and frequency. AI can adapt and learn from the feedback and responses of the victims and modify its tactics accordingly. This can increase the success rate and impact of the cyberattacks while minimizing effort and cost. Real-life applications have already appeared, including a fake video of India's National Stock Exchange CEO in which he appears to provide financial investment advice.

THE RANGE OF TOOLS AVAILABLE TO CYBERCRIMINALS IS EXPANDING

A large number of AI code-generation tools have recently appeared, including WormGPT, FraudGPT, Wolf-GPT, and Predator. Although we do not yet know how effective and impactful these tools are — and the development trajectory of specific tools remains uncertain (WormGPT closed within two months of launching due to claimed negative publicity, and many similar products are suspected to be fake) — the range of tools available to cybercriminals is certainly expanding. "We expect over time as adoption and democratization of AI models continue, these trends will increase," warned an FBI official in a 2023 statement.

At the same time, organizations are struggling to hire and retain cybersecurity specialists. Gartner predicts that by 2025, lack of talent or human failure will be responsible for more than half of all significant cyber incidents. As AI implementation and the discrepancy between hackers and defenders accelerate, it will be even harder to close the cybersecurity gap. Many companies are already facing an uphill battle as they try to close the gap. Total spending due to cyber breaches is predicted to grow (US $10.5 trillion by 2025 versus $3 trillion in 2015 according to Cybersecurity Ventures), but cybersecurity budgets only increased 6% on average in 2023, compared to 17% in 2022, according to a study from IANS and Artico.

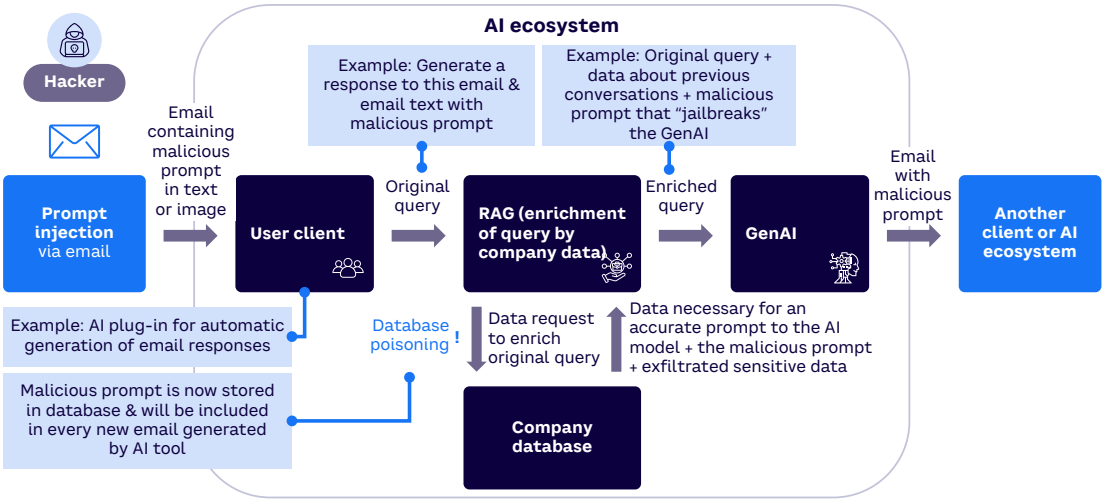
NEW AI ECOSYSTEMS INCREASE ORGANIZATIONS' ATTACK SURFACES

In addition to threats from direct AI weaponization, the increased integration of AI into corporate IT systems (and, eventually, operating technology) poses a significant danger. As companies accelerate the development of AI ecosystems, they are inadvertently creating new opportunities for hackers.

Gartner predicts that by 2026, more than 80% of enterprises will have used GenAI application programming interfaces (APIs) or models and/or deployed GenAI-enabled applications in production environments. This presents a concern for most organizations as it creates new vulnerabilities in their IT landscapes. However, companies have control over their AI integration approach, which means that implementing appropriate and proactive security measures to protect their IT environment is within their power.

Attacks on current GenAI models could lead to AI model theft, data manipulation, and/or data poisoning. Vendor rush and the deployment model used (e.g., which integrations are set up) could amplify this environment. Importantly, non-AI elements of the ecosystem (e.g., databases and applications) can often be accessed, increasing the potential damage. For example, in a test environment, researchers demonstrated that prompt injections are an effective method of hacking a database in an AI ecosystem (the AI model itself was not compromised). This highlights that the approach to AI integration is as important as the security of the model itself (see Figure 5).

Figure 5. Example of a GenAI ecosystem attack



RAG = Retrieval-augmented generation
Source: Arthur D. Little

To ensure safety, companies must develop AI systems using sound architectural principles. AI outputs should be verified, its actions should be controlled, and its access to sensitive data should be restricted. AI creates novel security challenges (e.g., data leakage), but fundamental security practices remain essential, including data protection, system visibility, and software updates.

HOW TO RESPOND

A solid cybersecurity program is more essential than ever to address emerging threats and ensure a strong basis for new technology implementation. The first line of defense against AI-based hacking is conventional cybersecurity measures. Defensive AI capabilities and tools can help counter hackers' improved efficiency and effectiveness (and the inevitable progress to the "transformation" wave of AI integration), but basic cybersecurity hygiene comes first. New security challenges can only be successfully met from a solid cybersecurity foundation.

THE FIRST LINE OF DEFENSE AGAINST AI-BASED HACKING IS CONVENTIONAL CYBERSECURITY MEASURES

In numerous conversations with business leaders, we discovered that although most believe their companies follow all cybersecurity best practices, a close look at their security teams usually reveals that's not the case. Figure 6 lists four examples of basic security questions. Unless you can provide a detailed answer to each, you should strongly consider reassessing your cybersecurity strategy.

For a more in-depth look at the cornerstones of cybersecurity strategy and the first steps toward comprehensively assessing your company's performance, see the ADL Viewpoint "Being Concerned Is Not Enough," which features a comprehensive, proven cybersecurity measurement framework.

Figure 6. Can you answer these basic cybersecurity questions?

4 KEY CYBERSECURITY QUESTIONS	INDICATIONS OF SIGNIFICANT PROBLEMS	GOOD ANSWERS
1. How many critical vulnerabilities do you have per server? What is the associated coverage of vulnerability scanning in IT & OT?	Don't know in detail/about 5 (coverage less than 95%)	Including month-to-month development, currently less than 0.3 per server with 99% coverage
2. When was the last time you conducted a board-level cybersecurity exercise?	Never/more than a year ago	2 within the last 12 months
3. What are your weakest cybersecurity areas/functions? What is the associated business risk in dollars?	Don't know/we don't charge	Domain XY (e.g., cryptography, especially for remote maintenance by provider X)/less than Y million dollars
4. After a ransomware attack, what proportion of your IT/OT data can be recovered within 24 hours/3 days after an attack?	Don't know/small proportion	24 hours: 85% 3 days: 98%

Source: Arthur D. Little

CEOs must ensure that cybersecurity is a high priority throughout the company, from top executives to part-time interns. Many business leaders mistakenly believe that only dedicated specialists can address cybersecurity, and the same myth exists for AI. The truth is, it's a strategic topic concerning the entire C-suite and board of directors. Security is not a direct source of business growth, but according to Sapio Research, 19% of business decision makers admit that a lack of cybersecurity credentials impacts their ability to win new businesses. Executives must make cybersecurity a strategic priority and highlight the AI challenge within it. Surprisingly, IT leadership involvement in AI security in this topic is limited: according to Gartner, although 93% of surveyed IT and security leaders said they are at least somewhat involved in their organization's GenAI security and risk management efforts, only 24% reported that they own this responsibility.

Employees are ultimately responsible for their own security, especially given the number of social engineering attacks. Training employees to be vigilant and aware of security fundamentals is a crucial part of protecting the company from threats.

DEVELOPING AI-BASED DEFENSES

The increased quantity and quality of AI-based attacks calls for AI-based solutions. AI can't replace humans, but humans can't replace AI, either. These solutions must include ways to securely integrate AI-based instruments into the corporate IT landscape.

Increasingly, companies and governments view AI as a serious cybersecurity enhancer and are starting to act. According to Gartner, 86% of decision makers believe the use of AI technology in cybersecurity tools will reduce the success of zero-day incidents (novel security events). IBM's survey of companies with more than 1,000 employees in more than 10 countries found that IT professionals believe more AI and automation in toolsets is the best way to improve threat-response times. Governments also expect value from AI — the US government recently announced a challenge with a $20 million prize fund in which hackers must compete using AI in cybersecurity.

Unlike malicious hacking, in which tools being released now mostly focus on reconnaissance/weaponization, defense tools currently being developed focus on a wider range of cybersecurity areas, such as attack prevention, infiltration detection, and threat processing. For example, AI tools can help prevent attacks by identifying vulnerabilities, managing access, and detecting phishing. When a breach occurs, AI can quickly detect it, automate response actions, and streamline incident handling for security teams.

We expect AI tools to provide support in the following areas:

- **Anomaly detection.** AI algorithms can continuously monitor network traffic and identify unusual patterns that could indicate a breach, spotting threats that traditional methods might miss. Anomaly detection of this type is still in its early stages, but when its full potential is achieved, it will effectively solve the zero-day vulnerability problem, currently the most powerful exploit.
- **Threat intelligence.** AI can analyze vast amounts of data to predict and identify emerging threats, helping organizations stay ahead of new malware, ransomware, and other threats.
- **Automated response.** AI can automate responses to detected threats, quickly isolating affected systems and blocking malicious activities, reducing attackers' window of opportunity.
- **Vulnerability management.** AI can scan and analyze networks and external perimeters for vulnerabilities, prioritizing them based on potential impact and suggesting the most suitable mitigation strategies.
- **Phishing detection.** Using natural language processing and ML, AI can identify and filter out phishing emails and other social engineering attacks before they reach end users.
- **Fraud detection.** AI systems can detect patterns indicative of fraudulent activities, such as unusual transaction volumes or anomalous credit card usage, and alert security personnel.
- **Security incident simulation.** AI can assess an organization's security posture by simulating attacks and predicting potential breaches, enabling proactive security enhancements. ADL has successfully applied this approach as part of cybersecurity assessments, incorporating AI-enabled reconnaissance and weaponization (penetration testing) stages. The tools employed were capable of pinpointing potential system vulnerabilities, with all high-confidence/high-warning and medium-confidence/high-warning alerts examined. To complete the assessment, human intervention was required to investigate false positives (basic alert checks were performed on those found to be potential issues).
- **Regulatory compliance.** Using scanning and vast data sets analysis capabilities, AI enables front-runners to better manage cybersecurity compliance. ADL has been leveraging this technology to evaluate compliance with International Electrotechnical Commission (IEC) 62443-4-1, IEC 62443-4-2, and EN 18031 standards. In our approach, we design questionnaires for the AI model to test conformity to these standards in an efficient and consistent manner. These tools help comprehensively address the cybersecurity compliance scope and assist in developing an accelerated plan to address the latest EU regulations. (An upcoming Viewpoint will explore optimizing efficiency with a combination of machine and human capabilities.)

AI-based solutions provide security benefits and make employees more efficient and effective. For example, AI-based defensive tools show promise in improving the effectiveness of cybersecurity specialists. Bugcrowd's survey of white hat hackers found that they acknowledged the importance of GenAI tools like ChatGPT in their work, with 21% stating that AI is already outperforming them in penetration testing.

Time-consuming threat investigation and false-positive identification are top issues for cybersecurity specialists, so we expect this to be a major development area for AI-based tools. IBM's study found that 46% of global respondents said their average time to detect and respond to a security incident increased over the past two years. Over 80% reported that manual investigation of threats slows down their overall threat-response times. Deep Instinct's research shows that false-positive–related work accounts for more than two working days of lost productivity per week. Gartner's interviews with nearly 50 security vendors found that, overwhelmingly, their first expectation from AI is to mitigate false positives.

The AI-enhanced tools already on the market generally fall under three groups:

1. **Phishing-detection tools** — use ML- or AI-based technology (e.g., IRONSCALES Themis, PhishML [part of PhishER], Hacker AI, and Cortex XDR)
2. **Full-defense tools** — aim to cover various stages of the cybersecurity process (e.g., Google Cloud AI and Darktrace's AI tools)
3. **AI-based penetration tools** — created for ethical hacking and optimizing the work of white hat hackers (e.g., Mayhem, DeepExploit, Pentoma, and Wallarm)

TIME-CONSUMING THREAT INVESTIGATION AND FALSE-POSITIVE IDENTIFICATION ARE TOP ISSUES FOR CYBERSECURITY SPECIALISTS

In the near future, we will likely see the application of AI both deeper and wider, covering most cyber domains except governance — narrow AI is not capable of replacing human problem-solving and decision-making capabilities. Because these tools will create vulnerabilities, they must be implemented carefully, applying all deployment best practices.

IT industry leaders are actively developing and sharing relevant expertise through publicly available frameworks, including Google's Secure AI Framework, Microsoft's Responsible AI Toolbox, Amazon's GenAI Security Scoping Matrix, and the Open Worldwide Application Security Project (OWASP) AI Checklist. These and other frameworks look at the problems from various angles, from core security guidelines to providing toolboxes for AI implementation. They discuss defending against AI-specific threats by implementing least-privilege control for AI tools, sanitizing inputs and outputs, and putting a human in the loop where feasible.

Robust governance structures must accompany this effort to ensure the responsible use of AI, comprehensive data security, and clear ownership of security processes. This combined approach will provide greater visibility into potential threats and granular control over security responses.

CONCLUSION

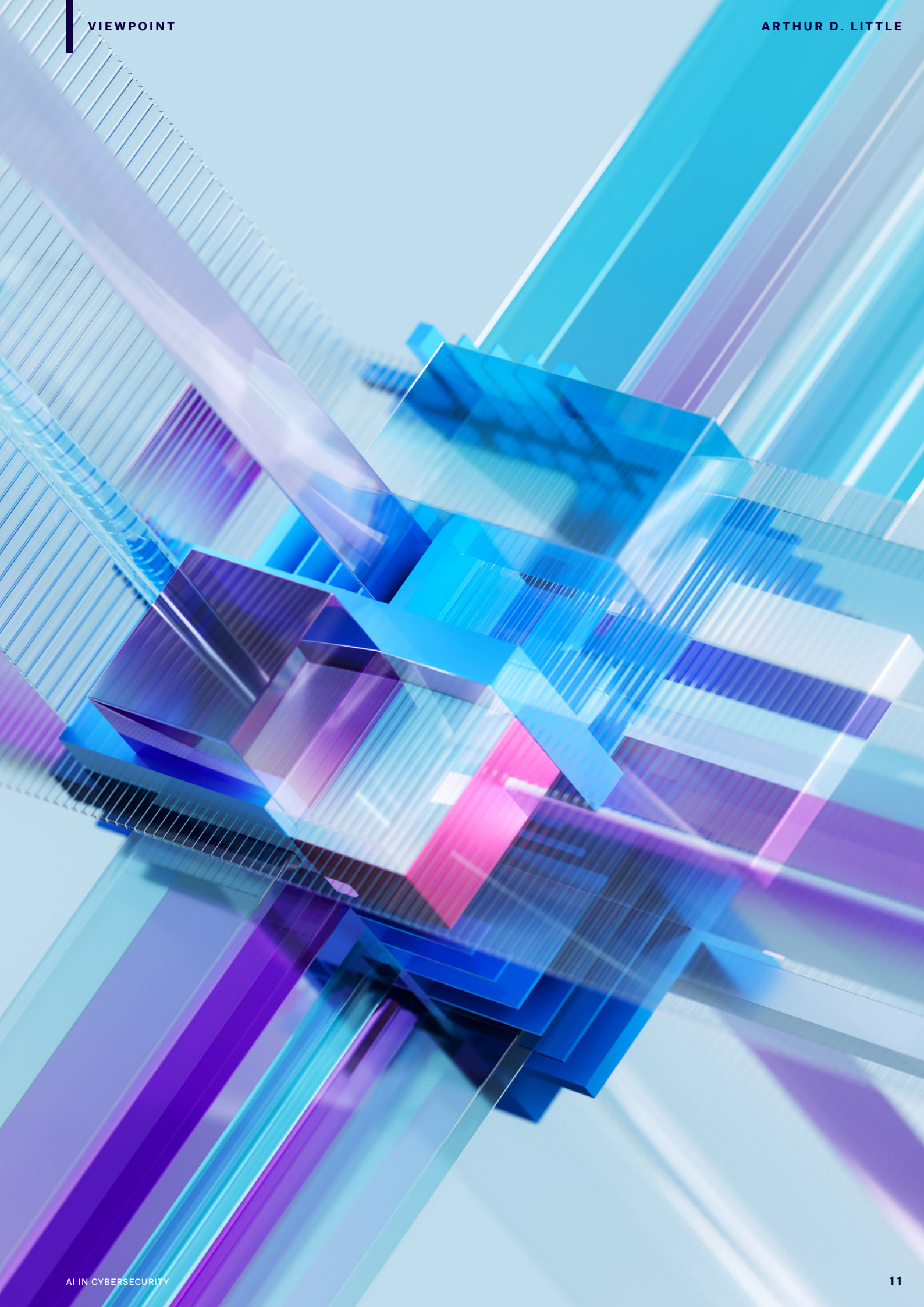
AVOID THE BLACK BOX EFFECT

AI-BASED DEFENSE TOOLS ACT AS ENHANCERS AND WILL NOT SHIELD YOU ON THEIR OWN

Unlike conventional cybersecurity measures, AI-based defense tools act as enhancers and will not shield you on their own. From an operational perspective, overreliance on AI-based defense tools can create a "black box" effect, in which business leaders blindly trust AI decisions without understanding the reasoning behind them, potentially leading to bad business and security decisions.

To defend against emerging cybersecurity threats, companies should:

1. **Evaluate their cybersecurity strategy,** starting with ensuring that a series of basic security questions can be answered in detail.
2. **Make cybersecurity a strategic priority,** involving senior leaders and the entire board.
3. **Leverage AI cyber-defense tools,** being careful not to expose the organization to additional threats. All new tools come with potential threats and should be integrated using stringent security measures.
4. **Develop AI capabilities to combat AI-based attacks,** recognizing that AI can augment, but not replace, human effort. Employees are the gatekeepers to your company's security, so they must be carefully trained to be vigilant about security fundamentals.



Arthur D. Little has been at the forefront of innovation since 1886. We are an acknowledged thought leader in linking strategy, innovation and transformation in technology-intensive and converging industries. We navigate our clients through changing business ecosystems to uncover new growth opportunities. We enable our clients to build innovation capabilities and transform their organizations.

Our consultants have strong practical industry experience combined with excellent knowledge of key trends and dynamics. ADL is present in the most important business centers around the world. We are proud to serve most of the Fortune 1000 companies, in addition to other leading firms and public sector organizations.

For further information, please visit www.adlittle.com.

Copyright © Arthur D. Little – 2024. All rights reserved.