Bell Labs Consulting

AI and security: exploring threats and opportunities

Quick take

The convergence of cybersecurity and AI presents a complex challenge with three key aspects: securing AI itself, defending against AI-enabled attacks, and leveraging AI for enhanced security. Securing AI involves protecting its software, data, and communications from threats like adversarial attacks and model extraction. Defending against AI-enabled attacks necessitates robust security measures, including stronger encryption, zero-trust architectures, and advanced threat hunting techniques. AI can also be a powerful tool for defense, enhancing threat detection, anomaly identification, and incident response. A comprehensive approach to AI security requires a vision of the future state and the application of formal risk management to identify threats, risks, and recommended countermeasures.

Introduction

Artificial intelligence (AI) is rapidly transforming various domains, including communications and other critical infrastructure sectors. With the proliferation of AI technologies comes the critical imperative to address security implications. The intersection of cybersecurity and artificial intelligence (AI) is a multifaceted topic with three key aspects to consider: securing AI itself, securing against attacks by AI, and securing through use of AI. Globally, regulations and guidelines are emerging related to trustworthy AI (e.g., the EU AI Act), articulating principles such as human oversight, privacy, transparency, fairness and technical robustness (of which security is a key aspect). By considering multiple dimensions of AI security, we aim to offer insights and recommendations to stakeholders involved in the design, implementation, use and regulation of AI systems.

Securing AI

Securing AI is a key aspect of the technical robustness of AI, which is one of the key principles of AI development addressed in the evolving space of AI regulations and guidelines globally. The security of AI systems involves protecting AI software (including platforms, application programming interfaces, algorithms and models) and data from unauthorized access, manipulation, or compromise. It includes protecting the confidentiality, integrity, and availability of the AI system and associated data.

Since AI is fundamentally a software construct, all existing threats against software can be relevant, including supply chain threats and malware. AI software should be developed using security-by-design principles, secure software development practices and adherence to the principle of “shift-left security”. In the latter, security is addressed proactively from earlier stages in the development lifecycle (for example, using DevSecOps, which is of particular relevance when AI software is developed to be cloud-native). Continuous vulnerability management (including scanning and patching) and proactive threat

hunting measures are required to protect AI systems in deployment. Since AI systems will require system administration, administrative access control should employ zero-trust principles, including strong authentication, fine-grained authorization and monitoring of accesses, ideally employing AI-enabled security analytics to detect anomalous activity by admin users. Critical AI systems should ideally be deployed on trusted computing infrastructure for integrity protection of critical software layers including hosting infrastructure (firmware, host OS, hypervisor, container engine, etc.) and the application layer itself.

Since AI systems typically process voluminous, complex and potentially sensitive data, all existing threats against data can be relevant, including threats against confidentiality, integrity, privacy and availability. Data security measures are required, including access controls, encryption, integrity protections and secure backups. Where encryption is employed, algorithms and key sizes should be quantum-safe for future-proofing against the potential breaking of cryptography through use of quantum computing.

Since AI systems are typically networked and expose interfaces including application programming interfaces (APIs), communications security and interface security are relevant. AI systems should be protected against common attacks such as denial-of-service (DoS), man-in-the-middle (MITM) and eavesdropping. Mutual authentication of endpoints for critical interfaces, use of encryption for sensitive communications and authorization for API requests are recommended. These latter aspects align with the zero-trust principle mentioned earlier in the context of AI system operation and administration.

Beyond the above measures to protect software, data and communications, AI systems are potentially subject to domain-specific threats against models and algorithms. These include adversarial attacks, where malicious actors attempt to manipulate model inputs to cause misclassification or erroneous outputs. Other AI-specific attacks include model inversion and model extraction.

When considering how to secure AI, we must recognize that there are multiple types of AI. For example, generative AIs (ChatGPT for example) are a type of machine learning (ML) AI, which in turn are a type of connectionist AI. There are other types of AI, including – for example – genetic algorithms or symbolic reasoning. Even within ML versions of AI, there are various types, including supervised, unsupervised, reinforcement and transfer learning. Deep learning (DL), generative AI (GAI), generative adversarial models (GANs) and large language models (LLMs) are all important variants of ML, suitable for a wide range of use cases including natural language processing, image processing and generation, and autonomous vehicles, to name but a few. AI research continues to accelerate, so there is not yet one canonical taxonomy. However, when securing AI, it is imperative to conduct threat analysis and risk assessment specifically for the type of AI in question, and for the use cases that it will support.

Securing against attacks by AI

As AI continues to advance, there is a growing concern about the potential use of AI in cyberattacks. The misuse of AI for malicious purposes, such as automated attacks or misinformation campaigns, introduces a unique set of challenges. Examples include AI-generated malware, AI-enhanced social engineering, automated reconnaissance, vulnerability scanning, and vulnerability exploitation. AI may potentially be used in crypto-analysis for cracking encryption. AI may enable complex, intelligent, multi-layer distributed denial of service (DDoS) attacks. Essentially, all steps in the cyber kill-chain can potentially be enabled and enhanced through use of AI.

All aspects of security should be bolstered to protect against attacks by AI, including people, process and technology aspects. Specific measures include stronger encryption, zero-trust architectures, and enhancements to risk, threat, access, vulnerability and incident management. Use of deception techniques

– honeypots, for example – can help frustrate attacks by AI and can also help with detection. Elevated maturity of vulnerability management is foundational to limit potential vulnerabilities that AI-enabled attackers can find and exploit. Advanced threat hunting techniques can help detect attacks and even breaches that may escape intrusion detection and monitoring systems.

Securing through use of AI

The utilization of AI for defense represents an opportunity to significantly enhance cybersecurity measures and will be of particular relevance in the context of AI-enabled attacks. AI-powered technologies can assist in threat detection, anomaly identification, and incident response, thereby bolstering the effectiveness of detective security measures. However, striking the balance between AI-enabled defenses and human expertise is crucial to address ethical concerns and avoid potential overreliance on automated decision-making.

Security monitoring yields massive amounts of potentially pertinent security data, which is difficult to analyze, whether in real-time or retrospectively for forensic analysis. AI can help with monitoring and detection, including anomaly detection. AI may even offer potential for predictive analysis to help with decision-making in the security context. Ultimately, AI will be employed for analytics-enabled security automation (security autonomics) to automate dynamic deception techniques, real-time defense, automated incident management, etc.

Beyond enhancing detection capabilities, AI can also enhance responsive capabilities to mitigate security attacks in real-time. Automation of defensive and responsive security measures can be enabled by AI. AI can act in milliseconds, significantly reducing the time between threat detection and response, which is crucial in mitigating the impact of fast-moving cyber attacks, particularly if these attacks are themselves orchestrated or enabled by AI.

Bell Labs Consulting approach

The BLC approach to security and AI is based on a formal threat analysis and risk assessment methodology augmented with a mature cybersecurity knowledge base. This risk-based approach ensures that potential risks and required countermeasures are comprehensively identified and properly assessed, leading to optimum balance of security investment against risks. BLC take a future-back perspective, starting with a vision of the future state of AI and security, enabling road-mapping of initiatives and investments to bridge between the present mode of operation and the desired future state.

Conclusion

Addressing the intersection of cybersecurity and AI requires a vision of the future state and the application of formal risk management to comprehensively identify threats, risks and recommended security countermeasures. Securing AI itself will involve continued application of current best-practices for software, data and network security, augmented by informed focus on AI-specific security issues relating to the various types of AI.

For further information please contact us at info.query@bell-labs-consulting.com

Learn more about Bell Labs Consulting at https://www.bell-labs.com/consulting/

About Nokia

At Nokia, we create technology that helps the world act together.

As a B2B technology innovation leader, we are pioneering networks that sense, think and act by leveraging our work across mobile, fixed and cloud networks. In addition, we create value with intellectual property and long-term research, led by the award-winning Nokia Bell Labs.

Service providers, enterprises and partners worldwide trust Nokia to deliver secure, reliable and sustainable networks today – and work with us to create the digital services and applications of the future.

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: 1051103 (September) CID214197