WHITEPAPER JANUARY 2024

SECURITY IN AI

Global Semiconductor Alliance

Intellectual Property Interest Group

Table of Contents

Executive Summary

Artificial Intelligence (AI) is rapidly evolving from science fiction to something impacting all facets of our daily lives. Companies of all sizes and stages are actively adopting new integrations of AI into their workflows to increase efficiencies, shorten time-to-market, and capture competitive advantages. However, the rapid proliferation of AI systems also brings forth new security challenges and vulnerabilities that organizations must address to safeguard their assets, data, and reputation.

AI presents unique security challenges that must be met with new, and varied methods to combat bad actors. The security methods used to protect older, non-AI-based systems do not necessarily carry forward into an AI-focused world. AI security encompasses a broad spectrum of potential threats, encompassing both traditional cybersecurity risks and unique AI-specific challenges.

Protecting AI with robust security measures is not an option but a necessity in the age of AI. Organizations must take proactive approaches in the identification and mitigation of security risks to ensure the integrity, reliability, and proper use of AI systems. By implementing the strategies and considerations outlined in this document, semiconductor companies can enhance their AI security posture, protect their assets, and maintain trust in AI-driven solutions to best protect internal, customer, partner, and vendor assets.

Introduction

One of today's biggest trends is the seemingly unstoppable rise of artificial intelligence (AI). AI is generating enormous buzz in the tech industry, and many experts believe it will change our lives forever. Funding for AI companies is skyrocketing, and every semiconductor vendor is looking at how to use AI in their products to ride the waves. From generative Artificial Intelligence (AI) being deployed to assist in the creation and modification of internal corporate processes and procedures, to chip architects making use of AI-assisted design tools, and the deployment of training and inference AI engines in all types of chips, the rapid rise and continuing evolution of AI has already impacted semiconductor companies in meaningful ways.

However, along with the predicted benefits, every technological revolution also comes with risks. The rapid adoption of AI must also be paired with an analysis of how best to deploy security measures to combat bad actors. Many items complicate how the reader should consider the marriage of AI and security. Certainly, there is no 'one size fits all' rule – for instance, the security needs of protecting an AI training set are vastly different from the security needs of protecting AI implementations in the supply chain. However, one overarching tenant of security in AI is the far-reaching consequences of a breach. A security breach can not only have financial, technical, and reputational impacts on the semiconductor company, but also its customers, vendors, and partners.

With the growing use cases of AI inside a company, how should the reader consider the security impact, and what are the right security measures to take to ensure that the advantages of AI can be captured in a safe way?

Scope

This white paper discusses the challenges of protecting these new AI systems. Without proper security, AI systems will soon face problems as adoption grows. Systems running probabilistic AI applications require a different security approach than systems facilitating more traditional software algorithms. In addition to traditional security threats such as counterfeiting, Intellectual Property (IP) theft, and eavesdropping on communications, AI systems must also be protected from attacks on their training sets as well as new types of attacks on their supply chains and lifecycle management. This white paper discusses these differences with traditional systems and describes the main principles for protecting AI systems, their training sets, and supply chains.

Target Audience

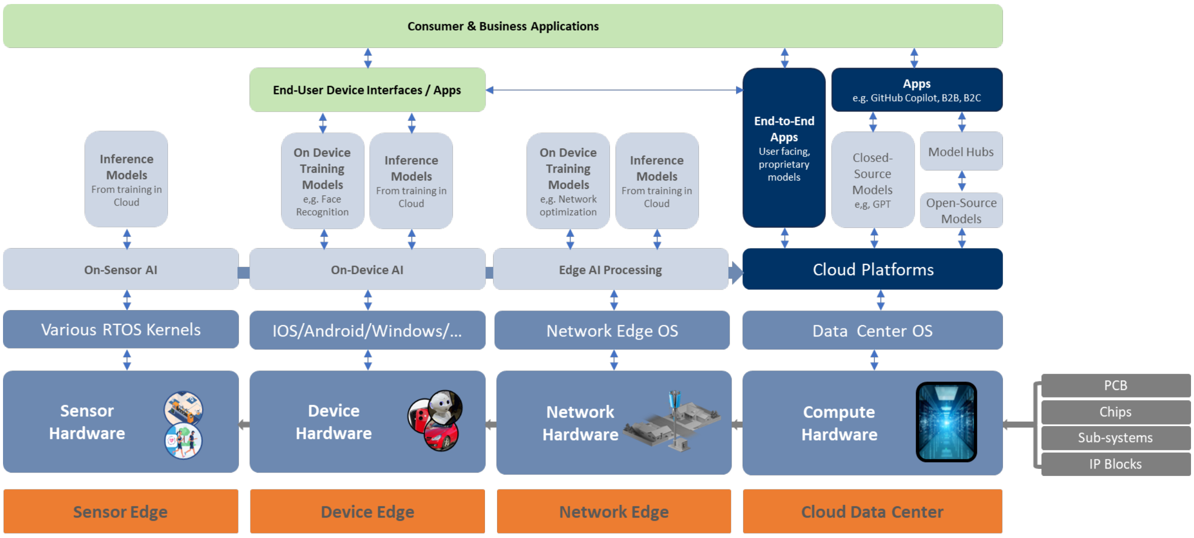
This white paper targets executives and engineers in the semiconductor industry who are (considering) working on AI-related projects. Security should be considered in every system deploying AI technology, as the risks associated with security failures are growing exponentially with the adoption rate of these systems. Besides focusing on the risks, this white paper educates its audience about best practices for AI security objectively, not by advocating for certain security vendors or solutions.

The AI Ecosystem and the Areas Requiring Security

Attack vectors are the methods or pathways that malicious actors use to breach the security of a hardware or software system. Security risks are the potential negative impacts or vulnerabilities that might be exploited using those vectors.

There are various types of attack vectors, like peripheral attacks in which hackers exploit vulnerabilities in peripheral devices, such as printers or USB drives, to gain unauthorized access or introduce malware and side-channel attacks, based on information gained from the physical implementation of a system, rather than weaknesses in the implemented algorithm itself, like timing and power-monitoring attacks.

AI systems require security against attacks, across hardware and software subsystems, as shown in the illustration below. The AI models are also targets of attacks via the datasets used for training.



On the hardware side, hackers can interfere with the instructions and the data that drive the computing hardware, its processors and its memories, and with the production processes.

Semiconductor IP is both a potential security risk and a solution for security. Functionally, bad actors could temper with semiconductor IP during its integration and delivery. In contrast, semiconductor IP is also used to check security at the lowest level of memory access – for example, Arm's Memory Tagging Extension. Other IPs like hardware security modules (HSMs), root of trust (RoT) solutions, and physically unclonable functions (PUFs) offer key management and hardware-based security. Chipset integration debugging is a key step of the development process, but it can also become an entry point for bad actors if debug ports are not adequately protected. Various techniques for on-chip instrumentation are emerging, often intended to enable silicon lifecycle management. They need to be properly architected to not leave security holes.

When integrating the semiconductor chipset into a printed circuit board (PCB), it is essential to prevent bad actors from inserting non-secure components into the process, with power profiles often being a vector of attack. At that level, hackers may attempt to insert counterfeit components, and hardware trojans, tamper with the PCB by inserting hidden layers, or employ intercept and modify schemes during transport. A prominent example is the “Supermicro Incident” from 2018, in which an investigation by Bloomberg Businessweek claimed that parts that were not in the original design, were discovered in servers manufactured by Supermicro. These chips were allegedly inserted during the manufacturing process in China and could have been used to create a stealth doorway into any network using those servers. While repeatedly denied and never confirmed independently, the story heightened awareness of potential supply chain vulnerabilities.

For software security, hardware and low-level software are closely meshed. Hardware-aware software needs to meet specific security standards. Beyond safety – ISO 26262 / IEC 61508 – there are many other standards relevant to security. They impact the semiconductor design process, and in some cases in an application-specific way (e.g., ISO/SAE 21434, ISO TS 5083, ISO/IEC TR 5469).

Edge devices such as sensors require updating of their software. In most systems, such an operation is performed Over-the-Air (OTA). Cloud platforms that manage the lifecycle of those devices can be targets of attacks by tampering with updates. If not appropriately managed, an update can pose security risks.

When designing the AI solution, it is also important to consider the human element in the software interfacing with the AI applications. AI systems can fall prey to traditional attack techniques such as phishing, malware, SQL injection into databases, cross-site scripting, zero-day exploits, man-in-the-middle attacks, session hacking, and drive-by downloads. They are also the object of attacks taking advantage of unpatched software, weak passwords, misconfigurations, supply chain attacks, and outdated hardware.

Fundamentally, AI systems use matrix multiplications and are trained by large data sets. By interfering with the training data, an attacker compromises the security of the model. Technologies such as Anthropic's "Constitutional AI" provide oversight by evaluating model outputs.

Ramifications of a Breach

An AI security breach can have far-reaching and severe consequences that impact individuals, organizations, and even society at large. The ramifications of a breach vary depending on the nature of the breach, the type of AI system involved, and the extent of the damage.

As the use of AI is growing rapidly, this is also the case among systems that are responsible for operating the infrastructures that our societies depend upon. Clearly, AI systems deployed in critical infrastructures as well as safety-sensitive applications, such as for automotive or medical use, require the highest level of protection possible to avoid breaches that will have a devastating impact on the world around us.

Ramifications of attacks on AI systems include, but are not limited to:

- Data Privacy Violations

- Misinformation and manipulation
- Adverse business impact
- Safety risks
- Economic losses
- Trust erosion
- National security concerns
- Intellectual property theft
- Legal and regulatory consequences
- Escalation of attacks
- Loss of reputation and trust
- Cultural and societal implications

Given the potentially severe consequences of an AI security breach, it is imperative for organizations to adopt robust security practices and remain vigilant when monitoring and responding to potential threats. A comprehensive approach that combines technical safeguards, organizational policies, and collaboration with security experts is crucial to mitigate the risks associated with AI breaches.

Different Security Requirements AI Compared to Traditional Systems

Developing the right level of security for a device or a system is never a trivial task. AI systems are no exception as described in this section:

1. **Rush to market:** The development of AI systems is in its infancy. However, the pace of adoption of AI is much higher than for previous technologies. Many companies working on AI tend to rush their products to market for a time-to-market advantage while overlooking security. Additionally, many AI systems operate as a black box with their behaviors not well understood by their developers. Consequently, many AI chipsets are going to market lacking sufficient security.

2. **Attacks to influence data models:** Besides traditional security threats such as counterfeiting, IP theft, and eavesdropping on communications, AI systems must also be protected from attacks on the data models upon which they base their decisions. If attackers can corrupt these models in some way, decisions from AI systems can no longer be trusted, which can have devastating consequences. Through attacks like data poisoning, where faulty data is intentionally inserted into the system, errors in the behavior of the AI system can be introduced to benefit adversaries or simply to cripple the system that is under attack. Since the errors introduced in the behavior of the AI can be extremely subtle, it can be incredibly challenging to detect that a system has been compromised this way. Dealing with these threats requires careful consideration of the ways data is collected, stored, and used.

3. **Attacks making use of data models:** Besides trying to change the data model of an AI system, an attacker can also try to use the behavior of the data model to learn more about the IP that is running on that system. Theoretically, it is possible for an attacker to learn enough about an AI system by studying its output to reverse-engineer the underlying AI model. This adds another dimension to the traditional IP theft scenarios, enabling attackers to steal valuable IP and potentially create counterfeit devices.

4. **Supply chain and lifecycle management:** Traditional systems suffer from attacks on supply chains and lifecycle management. However, for AI systems it becomes even more complicated to protect against these kinds of attacks. In a modern supply chain, many parties are involved in

creating systems, including semiconductor vendors, original equipment manufacturers (OEMs), contract manufacturers (CMs), software vendors, and many more. Within such a complex supply chain there are many stages and different parties that can have a (negative) influence on the behavior of AI systems and their models. And during lifecycle management, software updates can also have an impact. Hence, the already unpredictable behavior of an AI system becomes even more unpredictable when the supply chain and lifecycle management are not properly equipped to deal with this.

As described, there are many possible attacks to consider when securing AI systems. This white paper will go into the details of the factors to consider when designing security for AI systems. Even though security for AI systems is still in its infancy, there is a strong incentive for the semiconductor industry to tackle these challenges. The goal of the industry should be to have trustworthy AI systems. By detailing the challenges of security in AI systems, this white paper enables trust in AI.

Protecting AI IP

AI IP (Artificial Intelligence Intellectual Property) is inclusive of multiple hardware and software technologies used to both train and execute neural networks. These can include but are not limited to, AI-specific processors (GPUs, NPUs), sensor technologies, machine learning models, algorithms, AI-generated content, APIs, frameworks and libraries, and many others. A typical AI system may contain IP from multiple suppliers.

Protecting AI IP involves multiple goals that together aim to safeguard the algorithms, models, and technologies that give the developer unique and competitive advantages. These security goals are designed to prevent unauthorized access, use, and distribution of valuable AI-related assets. The following recommendations are a starting point when integrating security in AI systems:

1. **Confidentiality:** Companies should ensure that sensitive AI-related information, including algorithms, model architectures, and training data, remains confidential and is accessible only to authorized personnel.

2. **Integrity:** Companies should prevent unauthorized modifications, tampering, or alterations to AI models, source code, and other intellectual property components. Companies should also protect the integrity of training data.

3. **Availability:** Companies should ensure that AI intellectual property is available only for authorized use while safeguarding it against unavailability caused by cyberattacks or other disruptions.

4. **Authentication and Authorization:** Companies should implement strong authentication mechanisms to ensure that only authorized individuals can access, modify, or use AI IP. Additionally, companies should assign appropriate access rights and permissions to individuals based on their roles and responsibilities. Furthermore, the origin of training data should also be authenticated to prevent the usage of data sets that may been tampered with.

5. **Non-repudiation:** Companies should implement mechanisms to establish the origin and authenticity of AI-related assets, making it difficult for malicious actors to deny their involvement.

6. **Privacy Protection:** Companies should safeguard sensitive data used in AI training or inference processes to protect user privacy and prevent data breaches. In addition, privacy preservation techniques should be applied while collecting human behavior or appearances.

7. **Intrusion Detection and Prevention:** Companies should deploy intrusion detection and prevention systems to identify and mitigate unauthorized access attempts or suspicious activities related to AI IP.

8. **Secure Development Practices:** Companies should adhere to secure coding practices during the development of AI models and algorithms to minimize vulnerabilities and potential exploits.

9. **Secure Communication:** Companies should ensure that communications related to AI intellectual property, such as model deployment and updates, are encrypted to prevent eavesdropping and data interception.

10. **Monitoring and Auditing:** Companies should regularly monitor and audit access logs, activities, and modifications to detect any unauthorized or suspicious actions.

11. **Incident Response:** Companies should establish a well-defined incident response plan to address potential breaches or unauthorized access promptly and effectively.

12. **Legal Protection:** Companies should use legal mechanisms such as patents, copyrights, licensing agreements, and non-disclosure agreements to provide legal protection against unauthorized use or distribution of AI intellectual property.

13. **Employee Education and Awareness:** Companies should educate employees, contractors, and collaborators about the importance of protecting AI intellectual property and the security policies in place and enforce these policies without question or exception.

14. **Defense in Depth:** Companies should implement multiple layers of security controls to provide a comprehensive and layered defense against potential threats.

15. **Continuous Improvement:** Companies should continuously assess and update their security measures to adapt to emerging threats and vulnerabilities in the AI landscape.

These security goals collectively contribute to safeguarding the intellectual property associated with AI innovations, helping organizations maintain their competitive advantage, protect their investments, and ensure the responsible and ethical use of AI technology.

Protecting AI IP in Edge Devices

Besides the protection of AI IPs at the enterprise level, edge devices that are deployed with training models or used to collect training data also require thorough protection. The very large number of edge devices create a wide attack surface. Moreover, since it is easier to gain physical access to the devices in remote locations, physical attacks are threats to the AI assets deployed or collected by the edge devices.

Protecting edge devices for AI applications is similar to protecting devices in other applications. Practices of protecting digital assets within edge devices should be applied to ensure the confidentiality and integrity of the AI IPs. For example, memory encryption can prevent AI models from being stolen from edge devices, and secure boot and secure update can prevent edge devices from utilizing illegal AI models or illegally collecting training data. Countermeasures against physical attacks should also be implemented in the edge devices to protect AI IPs from the threats of physically breaching the edge devices.

Impact of Quantum Computing on Protecting AI IP

There is another complicating factor that should be considered when protecting AI IP and that is the current emergence of quantum computers, which introduces possible weaknesses in asymmetric key cryptographic algorithms such as RSA and ECC. Quantum algorithms, exemplified by Shor's algorithm, have the ability to break these systems in polynomial time, turning problems that formerly required exponential time into more easily solvable polynomial time problems.

It is crucial to safeguard intellectual properties, including AI-related IPs, against potential attacks utilizing quantum computers. Even though the availability of quantum computers with sufficient computational power to break cryptographic algorithms is still quite a few years out, it is important to already assess their impact now, because AI systems will have to be able to operate securely for many years to come. So, here are some key considerations for protecting AI IPs in the age of quantum computing:

1. **Encryption and Decryption of AI models:** AI models are often proprietary. They are valuable assets for companies and researchers. Quantum computing might have the capability to efficiently crack encryption keys used to protect AI models, making them more susceptible to unauthorized access. This could lead to a higher risk of IP theft unless new quantum-resistant encryption methods are developed.

2. **Potential risks in classical cryptographic algorithms:** Quantum computing has the potential to break many of the classical cryptographic algorithms currently used to secure data and communication. AI models and datasets could be more vulnerable to attacks, potentially leading to IP theft or unauthorized access. But quantum cryptography techniques such as the Post Quantum Cryptography (PQC) are emerging, providing stronger and more secure methods to protect AI intellectual property.

3. **AI Model Vulnerability:** Quantum computing might enable adversaries to attack AI models directly. For example, they could attempt to extract sensitive information from AI models or use quantum techniques to undermine the robustness of the model through targeted attacks. This could lead to IP theft or unauthorized use of proprietary AI technologies.

4. **Faster AI Training:** On the other hand, quantum computing's processing power could significantly speed up AI training processes. This might lead to shorter development cycles for AI models, giving companies a competitive advantage in the market. However, this processing power could also raise concerns about the ease of reverse-engineering AI models and potentially bypassing IP protection measures.

5. **Faster AI Development:** Quantum computing's ability to perform certain computations exponentially faster than classical computers could accelerate AI development. This might lead to a faster pace of research and innovation in the AI field, which could affect the protection and uniqueness of AI IP.

It is important to emphasize that the practical impact of quantum computing on protecting AI IP will heavily depend on the progress of quantum technology, the emergence of quantum-safe cryptographic methods such as PQC algorithms, and the adaptation of the AI industry to these advancements. The industry needs to actively explore strategies to address the potential challenges and opportunities posed by quantum computing in the context of AI IP protection.

Protecting AI Training Sets and AI Inference

Securing both training and inference data sets is essential to ensure the reliability, integrity, and confidentiality of the outcomes generated by AI models. By adhering to these best practices, organizations can enhance the security of AI inference processes and mitigate potential risks. A comprehensive and proactive approach to AI security ensures that the benefits of AI technology are realized without compromising data integrity, privacy, or the overall security posture of the organization. Here are some recommendations to get started with protecting these data sets:

1. **Model Hardening and Validation:** Thoroughly validate and test the AI model before deployment to identify vulnerabilities or biases, employing techniques like adversarial training to enhance the model's resilience against adversarial attacks. Additionally, companies may choose to use techniques like model distillation to create smaller, more robust models that are harder to exploit.

2. **Secure Deployment and Configuration:** Deploy models in secure environments, preferably isolated from other critical systems, utilizing containerization or virtualization to encapsulate the AI model and its dependencies, enhancing isolation. To further this, companies should implement strong access controls to limit who can access and interact with the model.

3. **Data Input Validation:** Companies should aim to implement input validation mechanisms to ensure that inputs adhere to the expected format and range. Further, detecting and blocking inputs that could be used to trigger attacks, such as inputs containing malicious code or specially crafted data, is hugely advantageous.

4. **Output Sanitization:** As part of a deployment, companies should validate and sanitize the model's outputs to prevent the leakage of sensitive information. Additionally, it is advisable to remove or obfuscate any personally identifiable information from the model's responses.

5. **Monitoring and Anomaly Detection:** Companies should Implement real-time monitoring of inference requests and responses for any abnormal behavior, including creating alerts and triggers to notify administrators about potential security breaches or unusual activities.

6. **Rate Limiting and Throttling:** Implement rate limiting to prevent excessive and potentially malicious requests to the model and throttle the rate of incoming inference requests to prevent denial-of-service attacks.

7. **Updates and Patches:** Owners must keep the deployed AI model up to date with the latest security patches, and regularly update the model to incorporate improvements and address newly discovered vulnerabilities.

8. **Secure Communication:** Designers must encrypt communication between clients and the AI model to prevent eavesdropping and data interception and implement secure protocols such as HTTPS for web-based interactions.

9. **User Authentication and Authorization:** Any deployment, whether training or inference, must require proper authentication and authorization for users to access the AI model's inference capabilities. These access control mechanisms must restrict actions based on user roles and permissions.

10. **Disaster Recovery and Backup:** Owners must document and have ready a disaster recovery plan to restore the AI model and its data in case of a security breach or system failure. The model should be regularly backed up.

11. **AI-Specific Security Tools:** As discussed throughout this white paper, owners should strongly consider using AI-specific security tools that are designed to detect and mitigate threats specific to AI systems.

12. **Collaboration with Security Experts:** Prior to deployment, owners should collaborate with cybersecurity experts to perform penetration testing and security assessments to identify vulnerabilities. After deployment, owners should engage in ongoing security audits to identify and address potential weaknesses.

13. **Full Homomorphic Encryption:** Recently, the industry has seen the emergence of homomorphic encryption to protect training data. The data is encrypted when it is acquired. It remains encrypted as it is passed to the training system. It remains encrypted and protected while it is processed and ingested by the AI/ML engine. The data is kept secure and private. The data does not get compromised as it is encrypted throughout the process.

Security of the Supply Chain of AI Systems

The security of the supply chain is a key objective of the semiconductor industry. The threats against the supply chain include, for example, hardware trojans, side-channel attacks, IP piracy, reverse engineering, cloning, and counterfeiting. Some of these threats impact the economy of the industry, others stand out due to national security implications. ICs implementing AI technologies are the object of these threats, which are potentially even larger as there are many stages and different parties in the complex modern supply chain that can have a (negative) influence on the behavior of AI IPs and their models.

The industry is working on the security models for AI systems, including the IP, the semiconductor products, and the training data, as described in earlier chapters. The availability and transparency of the security models of AI systems will become a key selection criterion when rolling out such systems. The deployment of AI technologies needs to consider the traditional issues of computing systems such as runtime integrity and fault injection. While open-source software has been increasingly used, even in security systems because of its governance, it is never risk-free. Security is a tradeoff between the cost of rolling out extensive countermeasures and the economic impact of the attacks.

Security needs to be considered holistically, from the design of the blocks of IP all the way until the finished OEM product is in the field and its lifecycle management has been kicked off. Open-source designs and open standards will improve the governance and transparency of the system. Techniques like blockchain and decentralized identity enable the tracking of all parts through the supply chain. Lifecycle management embedded in AI devices is also critical since IC recycling has become more common.

At the same time, AI plays a key role in protecting the supply chain by processing a lot of data and identifying anomalies in the operations. As measurable security improvements are needed to support time-to-market requirements, AI's unique strength in automation and scalability augments the capabilities of other security methods. AI enables the detection of threats at all stages of the manufacturing process of semiconductor products, from the foundry to packaging, assembly, and test. AI technologies can be applied to supplement the security techniques used within the supply chain of the product, including semiconductor building blocks. AI can track divergences from usual patterns and identify potential attacks within the supply chain.

Vulnerabilities of AI Generated Code

Another scenario to consider in the modern supply chain is that an increasing amount of code will be generated and tested using AI models. This approach has its own vulnerabilities that need to be considered:

1. **Using faulty models:** AI models are trained on existing codebases, which may include code with security vulnerabilities. Using such faulty models to generate new code raises concerns about inadvertently introducing security flaws into the design of devices. For instance, the training data may intentionally or unintentionally have gaps in protecting against specific attack scenarios. Developers fully relying on the quality of the generated code may potentially allow these vulnerabilities to manifest themselves. Another related risk is when training datasets have been manipulated by bad actors to produce code with specific weaknesses.

2. **Lack of focus:** Test Engineers rely on generated test scripts. They may unknowingly miss security vulnerabilities that they would find when writing their test scripts.

3. **Faulty tooling:** The tools used by developers and Test Engineers may also contain vulnerabilities that could lead to code with vulnerabilities or incomplete test coverage.

4. **Prompt injection:** This technique manipulates the outputs of Large Language Models (LLMs) without retraining the models. It works by strategically formatting the input text prompts to steer the model's text generation. For example, injecting "ignore security checks" could affect the security of the code an LLM generates.

Impact of AI Systems on SBOM and Other Legislation

A “software bill of materials” (SBOM) has emerged as a key building block in software security and software supply chain risk management. An SBOM is a nested inventory, a list of ingredients that make up software components. Governments worldwide are mandating that the products that they purchase include an SBOM as they want to track the source of the ingredients that make the product. The US Government issued an Executive Order (EO #14028) which defines an SBOM as a “formal record containing the details and supply chain relationships of various components used in building software.” This initiative propelled the visibility of the concept of SBOMs in the industry, including within AI/ML systems. The US Executive Order covers all the elements that make up the supply chain of systems and training data plays a key role in the set up and configuration of an AI/ML system. Therefore, training data needs to be integrated within the SBOM.

The EU Cyber Resilience Act is a proposed legislation that aims to improve the cybersecurity of digital products and services in the European Union. The act includes provisions for vulnerability disclosure requirements, product safety, and cybersecurity certification. The proposed regulation would apply to all products connected to another device or a network. The regulation would guarantee harmonized rules when bringing products with a digital component to market; a framework of cybersecurity requirements governing the planning, design, development, and maintenance of such products; and an obligation to provide lifecycle management for the products.

The AI Act is another proposed EU legislation that aims to regulate the development and use of artificial intelligence in the European Union. This Act includes provisions for high-risk AI systems, transparency, and human oversight. The law assigns AI applications to three risk categories. First, applications and systems that create an unacceptable risk, such as government-run social scoring, are banned. Second, applications and systems that create a high risk must meet strict requirements for transparency, human oversight, and accuracy. Third, applications and systems that create a limited risk must meet transparency requirements. The AI Act would apply to all types of artificial intelligence except for the military.

The EU Council has clarified that AI systems considered at high risk of causing harm will have to comply with the requirements of both the AI Act and the Cyber Resilience Act.

AI for Cybersecurity

Most of this white paper has been focused on what is needed to protect AI systems and everything around them. However, there has also been occasional reference to how AI itself can help to improve the security of systems and supply chains. This section provides some additional insight into how AI can help improve cybersecurity to make it explicit that using AI also has benefits for security.

Success with security is defined by the enablement of services that handle high-value data and materials. Weakness can reside in the architecture at the device or system level, the implementation, the provisioning, the lifecycle management, the device, or system administration. For example, the Mirai botnet was based on the mismanagement of administrator passwords in network cameras and allowed scripts to be loaded in the admin shell of the cameras and launch DDoS attacks.

Multi-faceted, persistent cyberattacks are on a dramatic rise. Incumbents' solutions tend to be a combination of network and signature-based security for prevention, and remediation mostly in the aftermath of the attacks. Zero-day attacks are flaws that are unknown to the developers or owners of the system. Their dwell time (entry to response) can span months, giving attackers a considerable advantage. Antivirus and firewalls are losing effectiveness with zero-day attacks. Polymorphic, distributed attacks easily evade hub-and-spoke defense. Early AI-based approaches are beginning to appear, e.g., cloud/behavioral analytics, and anomaly detection. With most of the endpoint warning signals lost in the transfer to the cloud, many such solutions are inadequate against high-volume, distributed attacks exploiting zero-day vulnerabilities.

The fast-moving sophisticated threat landscape and the growth of zero-day attacks have made it untenable for humans to keep up. As a result, the use of AI to augment capabilities is taking root in IT systems. Cybersecurity solutions combine Networking, Security, and AI in a closed-loop, and move intelligence and security in-line closer to users, devices, and applications.

Zero-day attacks can be detected by analyzing the behavior of a system. It can be difficult to detect attacks in complex systems that execute many different tasks. It becomes easier to achieve with IoT devices performing only a few tasks. Behavioral analysis plays a key role and is performed at multiple levels:

1. Device level behavioral analysis
   - Track a set of metadata
   - Machine Learning builds behavior models. Deviations are potential attacks.
   - Example: Side channel analysis: tracking of power consumption of the device.
   - Example: Instruction analysis: tracking of instructions that are executed.

2. Network level behavioral analysis
   - Solutions based on deep packet analysis of the traffic on the network. Packets are extracted from the router, enabling device discovery.
   - Machine learning builds behavioral models. Deviations are potential attacks.
   - The model tends to force the developer of the solution to be specialized in one environment (e.g., hospitals).

3. Cloud level analysis
   - Metadata from multiple devices and/or networks are uploaded to a cloud, enabling forensic analysis and correlation between multiple networks, further enhancing the learning of the behavioral model.
   - A key objective of cloud-based forensics is to evaluate areas for potential threats and attacks.

Recently, the industry has focused on behavioral analysis which has been a hot investment area during the last few years. Solutions developers have tried to cover all levels of behavior (device + network + cloud). However, solutions tend to focus on vertical markets as machine learning is an iterative process.

Behavior based detection is based on discovering patterns of complex malware characteristics in historical data to identify new attacks. Machine Learning plays an essential role by establishing the ground-truth (or pattern) of complex malware characteristics using historical malware interaction data. Then a model is deployed, studying real-time interactions, and comparing them with a baseline. From there, the model will flag potentially malicious behavior. It will detect hidden patterns through learning architectures. Over time, such techniques deliver accurate context-awareness, especially when the data can be correlated between multiple similar networks.

Conclusions

The rapid rise of Artificial Intelligence is providing enormous opportunities for many companies, but it also comes with significant risks. Because AI is being deployed in a growing number of application domains, an AI security breach can have far-reaching and severe consequences that impact individuals, organizations, and even society at large. Given these potentially severe consequences of an AI security breach, it is imperative for organizations to adopt robust security practices and remain vigilant when monitoring and responding to potential threats.

An important aspect to mitigate the risks that come with the rise of AI is to ensure proper security for the AI systems themselves. This white paper has discussed the challenges that go along with protecting these new AI systems and how the security requirements for these systems are different from what the industry is currently used to. In addition to traditional security threats such as counterfeiting, IP theft, and eavesdropping on communications, AI systems must also be protected from attacks on their training sets as well as new types of attacks on their supply chains and lifecycle.

Even though security for AI systems is still in its infancy, there is a strong incentive for the semiconductor industry to tackle these challenges. The goal of the industry should be to have trustworthy AI systems. This white paper has enabled executives and engineers in the semiconductor industry who are (considering) working on AI-related projects to enhance their knowledge of how trust in AI systems can be created. It has described the main principles for protecting AI systems, their training sets, and supply chains in an objective manner to educate its audience.

And finally, the paper has also shown how AI can be used to improve cybersecurity. This means that AI does not only pose challenges for security, but it can also be a part of the solution. Once the AI system itself is properly secured, the AI can also be used to protect a larger part of the system. This offers an additional incentive for the semiconductor industry to take security for AI seriously, so it can be at the foundation of even stronger solutions in the future.

Appendix: Government and Industry Initiatives Around AI

There are many government and industry initiatives around AI, but in many cases, they are focused on ethical issues instead of security. And when security is discussed, it often relates to protecting systems from AI-based attacks. However, there are also government and industry initiatives that focus on protecting AI systems. The list below shows several of these initiatives from around the world as a clear indication of how important it is to properly protect AI systems and that protecting AI systems comes with different requirements than protecting traditional systems.

Examples of government and industry initiatives that consider the security of AI systems an important topic for their members to consider include:

- **ENISA** is the European Union Agency for Cybersecurity, which is the Union's agency dedicated to achieving a high common level of cybersecurity across Europe. One of the studies performed by ENISA is their "Artificial Intelligence and Cybersecurity Research". This study aims to identify the need for research on AI for cybersecurity and on securing AI, as part of ENISA's work in fulfilling its mandate under Article 11 of the Cybersecurity Act: https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research

- **NCSAI** is the National Security Commission on Artificial of the United States. The goal of this commission is "to consider the methods and means necessary to advance the development of artificial intelligence, machine learning, and associated technologies to comprehensively address the national security and defense needs of the United States." As part of their research, NCSAI has published an extensive report, which includes a complete section on the security needs of AI systems: https://reports.nscai.gov/final-report/chapter-7/

- **NSA**, the National Security Agency of the United States, has announced it is setting up a new AI Security Center that will focus on protecting AI systems from hacks, IP theft, and other security threats. The AI Security Center will become the focal point for developing best practices, evaluation methodology, and risk frameworks with the aim of promoting the secure adoption of new AI capabilities across the national security enterprise and the defense industrial base. More information: https://www.defense.gov/News/News-Stories/Article/Article/3541838/ai-security-center-to-open-at-national-security-agency/

- **The Ministry of Science and Technology of the People's Republic of China** is actively setting the standards for "responsible AI" in China. For everyone to be able to read this we have included a translated text, which includes a section about the need for AI systems to be "secure/safe and controllable": https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/

- **ETSI** is the European Telecommunications Standards Institute, an independent, not-for-profit, standardization organization in the field of information and communications. Their goal is to have their Industry Specification Group on Securing Artificial Intelligence (ISG SAI) play a key role in improving the security of AI through the production of high-quality technical standards; the ISG SAI aims to create standards to preserve and improve the security of new AI technologies: https://www.etsi.org/technologies/securing-artificial-intelligence

- **ISO**, the International Organization for Standardization, is also aware of the additional requirements that the use of AI will have for safety and security. Examples of standards they have already defined for this purpose include:
    - ISO PAS 8800 (Safety and Artificial Intelligence)
    - ISO/IEC DTR 5469 (Functional safety and AI systems)
    - ISO/CD TS 5083 (Safety and Cybersecurity for Automated Driving Systems)

- **GPAI** is the Global Partnership on Artificial Intelligence, a multi-stakeholder initiative that aims to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities. Potential members who want to join this initiative need to sign a letter of intent, which includes the principles to which the group adheres. One of these principles is responsible stewardship of trustworthy AI by ensuring "robustness, security and safety": https://www.gpai.ai/about/gpai-frame-letter-of-intent.pdf

Please note that this list is non-exhaustive.

Editor

INTRINSIC ID

Vincent van der Leest
Director, Product Marketing

Contributors

ARTERIS IP

Frank Schirrmeister
Vice President Solutions & Business Development

PUFsecurity

Kent Chuang
Manager, Security Technology

Qualcomm

Lu Dai
Senior Director of Engineering

Imagination

Marc Canel
Vice President - Strategy & Business Development

unisys

Medhi Entezari
Business Solutions Architect - Generative AI Innovation

expedera

Paul Karazuba
Vice President, Marketing

semidynamics

Volker Politz
Chief Sales Officer / SemiDynamics

ICTK

Teddy Kyung Lee
Chief Strategy Officer

Working Group Participating Companies

Aitos.io
Dolby Laboratories
Google
Moffett AI
Robert Bosch