eclypses®

WHITE PAPER

Securing AI Data Applications: Safeguarding Against AI-Driven Attacks and Protecting Source Data Integrity

www.eclypses.com

Document ID: ECLYPSES -1532395320-328
Version No.: 2.0

Contents

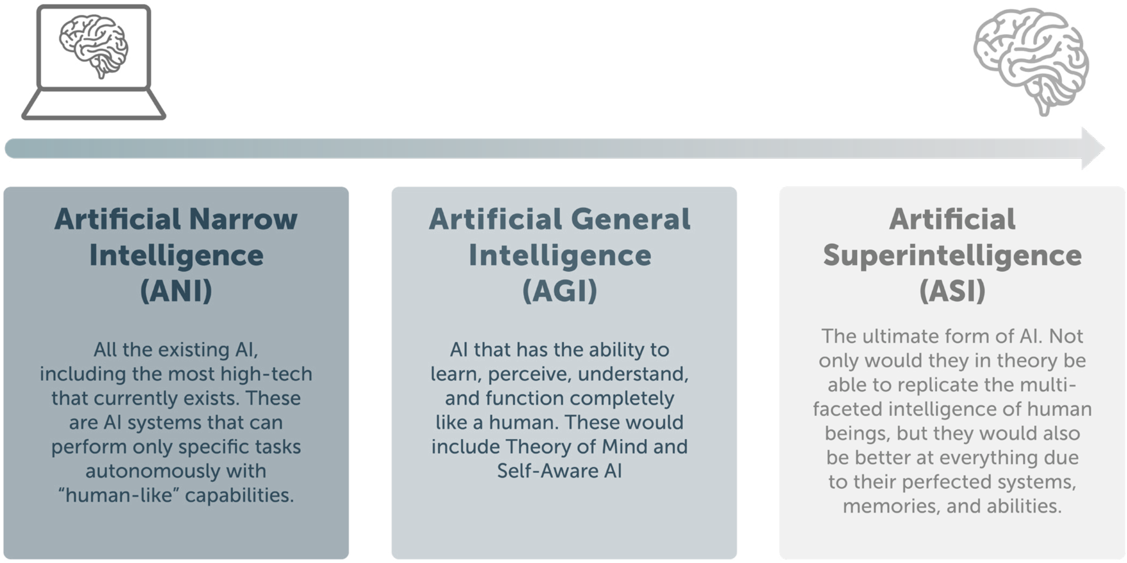
Introduction

The rapid adoption of AI technologies in various industries has opened new frontiers in innovation but has also introduced security concerns. AI-driven attacks, which leverage machine learning algorithms to exploit vulnerabilities, pose a significant threat to the confidentiality, integrity, and availability of sensitive data. Protecting AI applications and the underlying source data is crucial for maintaining trust and ensuring the responsible deployment of AI technologies.

Types of Artificial Intelligence

"Based on this criterion, there are two ways in which AI is generally classified. One type is based on classifying AI and AI-enabled machines based on their likeness to the human mind, and their ability to "think" and perhaps even "feel" like humans. According to this system of classification, there are four types of AI or AI-based systems: reactive machines, limited memory machines, theory of mind, and self-aware AI," according to a Forbes article.

- **Reactive Machines** - AI machines that can be used for automatically responding to a limited set or combination of inputs. They do not retain memory and therefore cannot learn from previous experience.
- **Limited Memory** - The most common type of AI for modern day AI applications, limited memory AI combines the functions of a reactive machine with the ability to learn from historical data to make decisions. Information and examples are fed into the AI, which then is able to create content or information derivative of it.
- **Theory of Mind** - A next level AI system that researchers are currently working to develop. This type of AI would be able to perceive humans as individuals, understanding their individual needs, emotions, beliefs, and thought processes.
- **Self-Aware** - The final stage hypothetical stage of AI which would essentially function as a human brain does. Possessing the ability to learn from memories and experiences, this type of AI would possess self-awareness with its own unique thoughts, needs, opinions, and beliefs.



Artificial Narrow Intelligence (ANI)

All the existing AI, including the most high-tech that currently exists. These are AI systems that can perform only specific tasks autonomously with "human-like" capabilities.

Artificial General Intelligence (AGI)

AI that has the ability to learn, perceive, understand, and function completely like a human. These would include Theory of Mind and Self-Aware AI

Artificial Superintelligence (ASI)

The ultimate form of AI. Not only would they in theory be able to replicate the multi-faceted intelligence of human beings, but they would also be better at everything due to their perfected systems, memories, and abilities.

- **Artificial Narrow Intelligence (ANI)** - All the existing AI, including the most high-tech that currently exists. These are AI systems that can perform only specific tasks autonomously with “human-like” capabilities.

- **Artificial General Intelligence (AGI)** - AI that has the ability to learn, perceive, understand, and function completely like a human. These would include Theory of Mind and Self-Aware AI

- **Artificial Superintelligence (ASI)** - This would be the ultimate form of AI. Not only would they in theory be able to replicate the multi-faceted intelligence of human beings, but they would also be better at everything due to their perfected systems, memories, and abilities.

AI Vulnerabilities to Your Enterprise

The ability for AI to synthesize and analyze vast amounts of data to generate meaningful and actionable information is increasingly important within an enterprise. However, as important as the information’s usefulness, knowing that the sources of that data are quality and untampered with by bad actors is just as important. This is true even for internal systems that serve as input into AI aggregation systems, especially if the internal systems are hosted in a cloud or hybrid environment.

One example of this is includes capturing customer demographics and current trends. Ensuring that this information is accurate is key to achieving better targeted marketing, especially when correlated with existing purchasing patterns of specific customers. If this information is inaccurate, the data becomes useless, wasting company time and effort.

Another example is the tracking of financial instruments and the impact that current trends may have on their value for future recommendations or investments. If the incoming data is compromised, manipulated, or injected prior to its aggregation into meaningful information, this can lead to skewed or improper results. The incorrect conclusions can then compromise business objectives.

Finally, it is important to consider what the potential consequences of receiving inaccurate information for company direction would be. If actionable information is sent to decision makers whether it be internal management, or recommendations to consumers, it becomes even more imperative that the information is true and trusted. As this kind of information can be even more private, it is also important to ensure that it is protected from theft by adversaries that lurk in the middle of the transmission.

The old adage certainly reigns true “Garbage In - Garbage Out.” When altered data is synthesized into meaningful information it may not result in a trusted decision or desired outcome. Protecting that data from its source to its destination to ensure that it is tamper proof is a critical requirement if the information is to be useful.

Best Practices to Enhance Your Security Posture of AI Data Applications

Adaptability: AI systems can adapt and learn from new data, ensuring that security measures evolve alongside emerging threats.

Efficiency: Automation of threat detection and response processes allows security teams to focus on more complex tasks, improving overall efficiency.

Scalability: AI-driven security solutions can handle large volumes of data and scale as an organization grows, providing a robust defense against cyber threats in various contexts.

Continuous Monitoring: AI enables continuous monitoring of systems, detecting anomalies in real-time and reducing the time window for potential threats to exploit vulnerabilities.

Things to Consider when Enhancing Security Posture of AI Data Applications

To ensure data that serves as input into an AI system is trustworthy, consider the following.

- Know the source: Ensure that the source of that data is trusted.
- Protect it in transit: Transmit the information over a tamper-proof channel.
- Protect it between internal systems: Ensure that data between systems is not compromised by internal threat actors.

To ensure information that is sent to decision makers as output from an AI system is trustworthy and secure, consider the following.

- Know the destination: Ensure that only intended audiences can view or act on the information.
- Protect it in transit: Transmit the information over a tamper-proof channel.
- Prevent unauthorized use: Ensure that the information can only be ingested by known and trusted recipients and that it has no value to any party that may have intercepted it.

Why Secure MTE Technology is the Solution

Cost-effective and potentially malicious, AI-driven attacks call for the best potential cyber-security implemented in your systems. With Secure MTE by Eclypses, the threat of these kinds of attacks are eliminated through real-time, unwavering data security across various application channels on prem and in the cloud. Secure MTE is the only proactive application centric solution that completely eliminates this growing risk, with Secure MTE the application does not waste any resources or time responding to non-validated traffic. Due to the validated nature of Secure MTE any application data, credentials, PII are also mitigated from this attack vector.

“Amidst AI’s evolving prowess in cyber warfare, fortifying our digital systems becomes paramount. Proactive defense measures, such as Secure MTE by Eclypses, not only safeguard against current threats like GPT-4 but also future AI advancements. In this ever-intensifying battle, vigilance and adaptability are our strongest shields,” says Jefferson Dance, Chief Innovation Officer at Eclypses.

In conclusion, integrating AI into data security strategies represents a significant leap forward in the ongoing battle against cyber threats. Leveraging the adaptability, efficiency, and scalability of AI-driven security solutions can empower organizations to create robust defense mechanisms. However, it is equally important to address the challenges associated with AI and ensure a balanced approach that prioritizes both security and ethical considerations. Ensuring that information is properly secured in a way that ensures data output can be trusted is key to utilizing AI to is maximum capacity safely. As we navigate the complexities of the digital age, the synergy between human expertise and artificial intelligence will be the key to maintaining a secure and resilient cyberspace.

Reach out to our team to learn more: contact@eclypses.com.