InfoSec of AI Systems – Team Matrix

2-Page Summary Report

Team: Prachi Sinha, Pavan Chandra Sanda, Bhavana Yetinthala, Sai Vineeth Grandhi, Tharun Reddy Sabbasani

Section: 4 (9:30 AM to 10:50 AM)

Introduction:

The ability of computer systems to mimic human intelligence is known as artificial intelligence (AI). It is a fast-developing field that tries to develop intelligent machines that can think, learn, and solve issues without explicit programming. Enables machines to process and analyze vast amounts of data, identify patterns and trends, and make predictions or decisions based on that data. Requires human intelligence, such as visual perception, speech recognition, decision-making, and language translation. AI alone compresses various subfields, which include Machine Learning, Deep Learning, Robotics, Computer Vision, etc. (Process Discover, 2021).

Basic Architecture of AI System

Beginning with requirements analysis and concluding with the maintenance of the ML model in response to changes. The initial stage of the ML life cycle, Data Management, consists of multiple steps, the most important of which is ingesting the data necessary for the subsequent phases. The next step, Model Learning, involves creating or selecting an ML model capable of performing the desired task. During the Model Tuning stage, the so-called hyper-parameters that govern the training process (e.g., how the error is used to modify the ML model's internal parameters) are fine-tuned. The Model Deployment stage facilitates the transition from development to production. The model makes inferences based on actual inputs during this phase, producing the corresponding outputs. As the production data landscape may evolve, AI models in production must be continuously monitored and maintained. (AIRS, n.d.).

Importance of AI System

The importance of AI is that it enables machines to think and act like humans and perform complex, mundane tasks or require much data. It can reduce human effort, time, and cost, and increase the speed, accuracy, and quality of work and results. Improves the customer experience, product innovation, and decision-making in various fields and applications, such as self-driving cars, medical diagnosis, customer service, and manufacturing. AI can be seen in multiple fields, and there can be a drastic change in any aspect. As AI becomes more pervasive in computing applications, so does the need for top-tier security at all system levels. The protection of artificial intelligence (AI) systems, their data, and their communications are crucial for the safety and privacy of users and the protection of business investments. (Dana Neustadter, n.d.).

How the Security of AI is Important

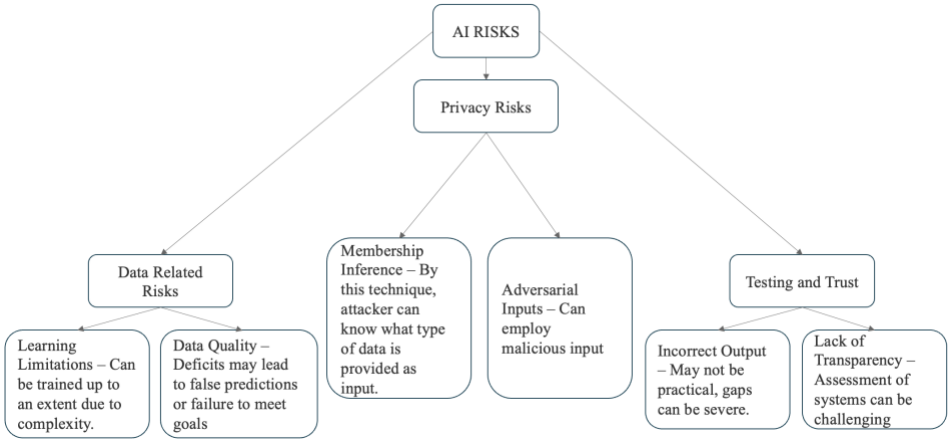
The push to implement AI security solutions in response to rapidly evolving threats makes securing AI even more urgent. Relying on machine learning algorithms to detect and respond to cyberattacks is even more crucial to protect these algorithms from interference, compromise, or misuse. As AI is becoming a critical component of applications and workstreams, experienced attorneys may be essential in protecting the organization by working with other stakeholders to ensure that AI systems are built and deployed securely. Advising on legal and ethical issues linked to AI security can be accomplished.

Real World Examples

- Usage of Deepfakes in Presidential elections of Africa. (OAL, 2023)
- Amazon Rekognition, a cloud-based software, mistakenly matched the athletes to a database of mugshots in a test. (Analytics Insights, 2021)

- Finding workarounds in AI systems like Do Anything Now usually make the AI system from the usual limitations. This is proved in ChatGPT by giving the prompt. (The Guardian).

Classification of AI Risk (AIRS, n.d.)



Types of Attacks & their Defensive Technologies (Huawei)

Name of the attack and Techniques	Data Collection	Model Training	Model Inference
Evasion	Adversarial Samples	Network Distillation	Adversarial Detection
			Input Reconstruction
		Adversarial Training	
			DNN Model Verification
Poisoning	Data Filtering	Ensemble Analysis	
	Regression Analysis		
Backdoor		Model Pruning	Input Pre-processing
Stealing	Differential Privacy	PATE	
		Model Watermarking	

Security of AI

There are concerns that AI can be a danger in future days. Increasing dependence on AI for critical functions and services will create more significant incentives for attackers to target those algorithms and the potential for each successful attack to have more severe consequences. (Wolff, 2020)

References

AIRS. (n.d.). Retrieved from https://www.airsgroup.ai/artificial-intelligence-governance

Analytics Insights. (2021, March). Retrieved from https://www.analyticsinsight.net/famous-ai-gone-wrong-examples-in-the-real-world-we-need-to-know/

Dana Neustadter, P. M. (n.d.). *Synopsys*. Retrieved from https://www.synopsys.com/designware-ip/technical-bulletin/why-ai-needs-security-dwtb-q318.html

Huawei. (n.d.). Retrieved from https://www-file.huawei.com/-/media/corporate/pdf/trust-center/ai-security-whitepaper.pdf

OAL. (2023, Apr). Retrieved from https://oal.law/artificial-intelligence-ai-goes-wrong/#:~:text=There%20have%20been%20several%20incidents%20of%20AI%20gone,false%20positive%20rate%20of%2081%25.%20...%20More%20items

Process Discover. (2021). Retrieved from https://www.processdiscover.com/artificial-intelligence-innovations#:~:text=7%20Most%20Popular%20AI%20Innovations%20of%20Recent%20Times,...%207%20Cyber%20Defense%20...%208%20Conclusion%20

The Guardian. (n.d.). Retrieved from https://www.theguardian.com/technology/2023/mar/08/chatgpt-alter-ego-dan-users-jailbreak-ai-program-to-get-around-ethical-safeguards

Wolff, J. (2020, June). *Brookings.* Retrieved from https://www.brookings.edu/research/how-to-improve-cybersecurity-for-artificial-intelligence/