The Cybersecurity Readiness Podcast Series

Episode Title	Securing AI's Blind Spots: The Hidden Risks in Enterprise AI Adoption
Podcast Series	The Cybersecurity Readiness Podcast Series https://www.cybersecurityreadinesspodcast.com/
Host and Producer	Dave Chatterjee, Ph.D. https://dchatte.com
Guest	Oliver Friedrichs, Founder and CEO, Pangea
Summary Pitch	The adoption of Artificial Intelligence (AI) and Generative Artificial Intelligence (Gen-AI) applications and tools are exploding. A McKinsey report estimates the adoption rate to be around 250% from 2017 to 2024. The global AI infrastructure market is projected to reach more than $96 billion by 2027. AI applications are being used to empower every organizational function and industry, from logistics and supply chain to manufacturing, healthcare, finance and banking, marketing and sales, and customer sales. However, such adoption and use of AI tools and platforms has greatly expanded the attack surfaces and the attack vectors. They are presenting many more opportunities for the hackers to break into systems and networks and also violate individual privacy and reputation, and thereby causing irreparable harm and damage. In this episode, Dr. Dave Chatterjee and Oliver Friedrichs, Founder and CEO, Pangea, discuss the risks associated with adoption and use of AI and Generative AI applications and platforms. They discuss examples of AI-powered attacks, such as prompt injection attacks, training data poisoning, deepfake scams, and ransomware attacks, and stresses the need for continuous learning and proactive security measures. They also underscore the importance of continuous security assessments, well-

	rehearsed incident response plans, robust checks and balances, and AI literacy for individuals and organizations.
Action Items and Discussion Highlights	▪ Be aware of the different types of AI risks such as the top ten list produced by OWASP. o Prompt Injection o Insecure Output Handling o Training Data Poisoning o Model Denial of Service o Supply Chain Vulnerabilities o Sensitive Information Disclosure o Insecure Plugin Design o Excessive Agency o Overreliance o Model Theft ▪ AI risk mitigation best practices include: o Continuous security assessments o Well-rehearsed incident response plan o Customized training o Leverage AI tools and technologies to proactively detect and thwart attacks o Implement process controls to prevent fraudulent transactions. ▪ Establish robust checks and balances, such as multi-party approvals for high-value transactions, to prevent fraudulent activities. ▪ Continuously stay up to date on the latest AI security research and developments to adapt security strategies accordingly. ▪ Encourage employees and individuals to become more AI-literate to better understand and manage the risks associated with AI adoption.

Time Stamps	00:00 – Introduction 01:58 – Guest's Professional Highlights 03:45 – Current state and evolutionary trends of cybersecurity tools and technologies 05:56 – Growth and use of AI and Gen AI 10:58 – AI Enabled Expanding Attack Surfaces 13:36 – The Hidden Risks 20:58 -- Role of humans in an AI-driven world 21:51 -- Securing the Internal AI chat bot platform of an educational institution 24:38 – AI-powered attacks 29:17 – Recommendations and Best Practices 38:03 – Closing Thoughts
Memorable Oliver Friedrichs Quotes/Statements	"I think effectiveness of security tools has improved dramatically, and our ability to respond quickly continues to improve dramatically, and that's powered even more by AI now." "AI is becoming this horizontal layer that's being embedded into almost everything as a capability to make everything better." "I would say the majority of developers aren't security experts; so just shoving a large language model into an application without really considering security is a really risky proposition, especially if you're working for an enterprise that has a customer facing interface to a large language model, that's where things become a real risk."

	“The attackers are creating new forms of prompt injection attacks regularly to mislead models.” “The only way to move forward is to embrace it and learn it and understand it and learn how to be the operator of AI and control it.” “Offensive use of LLMs and generative AI is a real thing, and it's not only allowing the attackers to scale more effectively in a way that there's really no limit in terms of being able to generate both the content, but also cyber-attacks, in particular, being able to navigate networks and actually create next generation worms and threats that can propagate independently and actually write code to replicate themselves in ways that are not detectable.”