The Cybersecurity Readiness Podcast Series

Episode Title	Securing Artificial Intelligence (AI) Applications
Podcast Series	The Cybersecurity Readiness Podcast Series https://www.dchatte.com/podcast/ https://the-cybersecurity-readi.captivate.fm/
Host and Producer	Dave Chatterjee, Ph.D. https://dchatte.com
Guest	Chris Sestito (Tito), Co-Founder and CEO of Hidden Layer
Summary Pitch	As artificial intelligence (AI) technologies continue to evolve and be leveraged, organizations need to make a concerted effort to safeguard their AI models and related data from different types of cyber-attacks and threats. Chris Sestito (Tito), Co-Founder and CEO of Hidden Layer, shares his thoughts and insights on the vulnerabilities of AI technologies and how best to secure AI applications.
Time Stamps	00:02 -- Introduction 01:48 -- Guest's Professional Highlights 03:55 -- AI is both a cure and a disease 04:49 -- Vulnerabilities of AI 07:01 -- Hallucination Abuse 10:27 -- Recommendations to secure AI applications 13:03 -- Identifying Reputable AI security experts 15:33 -- Getting Rid of AI Ethics Teams

	19:18 -- Top Management Involvement and Commitment
Memorable Chris Sestito Quotes/Statements	"Artificial intelligence systems are becoming single points of failure in some cases." "AI happens to be the fastest deployed and adopted technology we've ever seen. And that sort of imbalance of how vulnerable it is and how fast it's getting out into the world, into our hardware and software, is really concerning." "When I talk about artificial intelligence being vulnerable, it's vulnerable in a bunch of ways; it's vulnerable at a code level, it's vulnerable at inference time, or essentially, at real time when it's making decisions, It's vulnerable at the input and output stages with the users and customers and the public interacting with your models, it's vulnerable over networks, it's vulnerable at a generative level, such as writing vulnerable code." "Hallucination abuse would be the threat actor trying to manage and manipulate the scope of those hallucinations to basically curate desired outcomes." "We should be holding artificial intelligence to the same standards that we hold other technologies." "The last thing we want to do is slow down innovation, right? We want to be responsible here, but we don't want to stop advancing, especially when other entities that we can be competing against, whether that's in a corporate scenario, or a geopolitical one, we don't want to handcuff ourselves."

	"If we're providing inputs and outputs to our models to our customers, they're just as available to threat actors. And we need to see how they're interacting with them." "If you're bringing a pre trained model, and and you're going to further train it to your use case, scan it, use the solution to understand if there is code where it doesn't belong." "If we're providing inputs and outputs to our models to our customers, they're just as available to threat actors. And we need to see how they're interacting with them." "Red teaming models is a wonderful exercise, but we also need to look at things that are a little bit more foundational to security before we get all the way to AI red teaming." "The threats associated with artificial intelligence are the exact same threats that are associated with other technologies. And it's always people. It's always bad people who want to take advantage of the scenario and there's an enormous opportunity to do that right now."