Conclusion

<1 min

Through the past several exercises, we've seen a handful of ways AI can be attacked. While there is a great deal of research and content spent on studying how to defend against AI-related attacks, we need to consider the other side as well.

AI, at its core, is software. Software is written by people, and people are far from perfect. Because of this, AI systems can have vulnerabilities introduced into them during development. When developing AI, similar security considerations and practices should be made:

1. Teams should regularly review their codebase.
2. Teams should monitor how the application is trained, and they should perform regular security reviews. Failure to perform reviews could expose sensitive information or lead to unexpected severe behavior from the AI system.