Security Briefs 041 – AI Security Challenges Part 3

Operational Security and Manipulation

Welcome back. In this episode of our series on AI Security Challenges, we will talk about Operational Security.

AI doesn't just threaten information we put into it. It also creates new risks when adversaries use it against us.

One of the fastest-growing threats is phishing. AI can generate emails that are free of spelling mistakes, formatted professionally, and tailored to look like they came from someone you know. A team member receives an urgent message from "you," asking them to forward sensitive files. It looks convincing — but it's a trap.

Another danger is deepfakes — videos or audio clips that look and sound real but are entirely fabricated. These tools are already being used in scams. Imagine a video surfacing online of a pastor saying something inflammatory. The video is fake, but by the time the truth comes out, the damage is done.

In conflict zones, false news created by AI has already sparked fear and division. In one example, images of an explosion were generated by AI and circulated as real events, triggering panic before authorities could clarify.

The problem isn't just what AI can create — it's our tendency to trust it. If we lean on AI for crisis updates or quick decisions, we may be deceived. Operational security requires layered verification: check sources, confirm through trusted human channels, and resist the temptation to take AI-generated information at face value.

AI can multiply our reach — but it can also multiply our vulnerability.

Now you know.

In our next episode, we'll explore another angle: privacy and identity risks when personal information is misused by AI.

Episode Summary

This episode highlights operational risks of AI, focusing on phishing, deepfakes, and disinformation. Real-world examples show how AI-generated content spreads faster than corrections. We emphasize layered verification and resisting the temptation to outsource decisions to machines.