The Hague Centre for Strategic Studies

Stellenbosch University iYunivesithi Universiteit

Ministry of Foreign Affairs

AI Agents in the Wild: What Companies Are Getting Wrong About Security

DATA SYNC

Introduction

AI agents are no longer a future concept. They are being deployed inside companies across every sector right now, automating workflows, making decisions, and accessing sensitive systems with a level of autonomy that most security teams were never designed to handle.

The problem is not the technology. It is the gap between how fast organisations are deploying AI agents and how slowly they are thinking about what that means for security, governance, and risk. Real incidents are happening, and most companies are not yet aware of the attack surfaces they have opened.

In this challenge, you step into the role of a security consultant called in to assess a company that has deployed multiple AI agents across its operations. Your task is to identify what went wrong, explain why it matters, and propose how to fix it, before a breach makes the decision for them.

Start by exploring what AI agents are and how they differ from traditional automation or chatbots. What new attack surfaces do agentic systems introduce? What security principles are most commonly violated when organisations deploy AI agents? What does responsible AI agent governance look like in practice?

Tasks and Objectives

Your team will analyse a fictional but realistic scenario involving a financial services company that has deployed five AI agents across its operations. The company has experienced a series of security incidents. Your task is to diagnose what went wrong, assess the risk, and present a remediation strategy to the company board.

Key Tasks

Your 3-minute video proposal must cover (at least) the following:

1. The security failures you identified across the AI agent deployment and the root causes connecting them.
2. Your risk prioritisation of the incidents, with justification based on the regulatory and operational context of a financial services company operating in Southern Africa.
3. A practical 90-day remediation plan the board can act on, with specific and prioritised recommendations.

Final Requirement

Your plan should be simple, practical, and realistic, helping the company protect its systems while ensuring that operations remain safe and uninterrupted.

April 7th, 12:00PM CET/SAST is the deadline for you to hand in a video of max. 3min to cyberschool@hcss.nl, in which you explain your Challenge solution and proposal. This is the final submission moment and from this submission, a Challenge-winning team will be determined. More information on this will be communicated via email.

Contributor

Data Sync Consulting is a fractional CIO and Chief AI Officer advisory firm specialising in AI adoption, IT strategy, and digital transformation. Data Sync works with executive teams to help organisations adopt AI in ways that are practical, secure, and built to last, with a focus on leveraging existing systems rather than replacing them.