CYBERSECURITY 101 · eBuilder Security

Responsible AI in Security

AI is already inside most security teams, scoring alerts and drafting queries, but in most organisations adoption has run ahead of governance. Here is what responsible AI means in practice, and the one habit that keeps it safe.

WHAT IT IS

Responsible AI in security means using and building AI so it stays accountable, fair, transparent, secure and under meaningful human control. It has two sides, using AI responsibly inside your defences and securing the AI systems themselves against misuse and attack. Two years ago this was a research subject. Now it is a duty, because the EU AI Act sets binding rules and boards are personally accountable for security oversight in Sweden.

WHY IT MATTERS

€35 million or 7 percent

Maximum EU AI Act fine for using prohibited AI, of global annual turnover.

Regulation (EU) 2024/1689, 2024

2 August 2026

When most obligations for high-risk AI systems start to apply.

Regulation (EU) 2024/1689, 2024

Six

Non-existent court cases an AI invented in one US legal filing, all cited as real.

US District Court, SDNY, 2023

WHAT HAPPENED

In 2023 a New York lawyer used ChatGPT for legal research and filed a brief citing six court cases that did not exist. The model had invented them, confidently, with quotations and citations. The court sanctioned the lawyers and fined them 5,000 dollars. One habit would have caught it, checking each AI-supplied fact against the primary source before it left their hands. (US District Court, Southern District of New York, 2023)

RED FLAGS TO WATCH FOR

AI is confident even when it is wrong, so watch for the moments where that confidence goes unchecked.

- A confident AI answer accepted and acted on without anyone checking the source.
- Confidential code, data or documents pasted into a free public AI tool.
- AI tools in daily use that no one has approved, inventoried or reviewed.
- An automated decision that blocks, deletes or affects a person with no human sign-off.

THE ONE RULE THAT STOPS IT

Keep a person in the loop and verify what an AI produces against a trusted source before you act on it.

THEN BUILD THE HABIT AROUND IT

- Write a short AI acceptable-use policy naming approved tools and what data must never go in.
- Never paste confidential code or personal data into a public AI tool.
- Require human review for any AI decision that blocks, deletes or affects someone.
- Keep an inventory of every AI tool in use so shadow AI has nowhere to hide.

MYTHS AND FACTS

MYTH

Responsible AI is just an ethics nicety.

FACT

It is now a legal obligation. The EU AI Act sets binding rules and can fine prohibited AI up to €35 million or 7 percent of global annual turnover.

MYTH

If we buy an AI tool, responsibility sits with the vendor.

FACT

You remain accountable for the outcome. A Canadian tribunal held Air Canada liable for its own chatbot's incorrect advice in 2024.

MYTH

AI in the SOC removes human error.

FACT

It adds automation bias. Models can hallucinate and be manipulated, so a person must stay able to review and override consequential decisions.

THE LEGAL DUTY

Responsible AI is now a legal duty. The EU AI Act (Regulation (EU) 2024/1689) sets binding risk-based rules, and in Sweden the board is personally accountable for security oversight under Cybersäkerhetslagen, in force since 15 January 2026.

GET HELP

Responsible AI is mostly governance, oversight and training applied to a new kind of tool. Read the full eBuilder Security guide at the link below for the principles, the risks and the controls in full.

Read the Full Article →