AI guardrails AT A GLANCE

A practical control model for using AI in the workplace without leaving data, permissions, and user behaviour to chance

The Core Idea: Good guardrails don't stop useful AI. They make useful AI safer, clearer, and easier to trust

1. Discover

Before a business can govern AI, it needs visibility into what is already being used.

- Identify which AI apps employees are using now
- Separate sanctioned tools from unsanctioned ones
- Treat Shadow AI as a visibility issue first, not just a blocking issue

2. Decide

Set a clear position on which tools, data types, and use cases are acceptable.

- Define which AI tools are approved for work use
- Be explicit about what can and can't be pasted or uploaded
- Create simple rules staff can actually follow under pressure

3. Protect

Layer technical and data controls around the tools that are allowed.

- Tighten permissions and reduce oversharing in Microsoft 365
- Use data protection controls to stop sensitive information leaking out
- Assume prompts, files, an d external content can all create risk

4. Govern

Treat AI as an ongoing operating model, not a one-off release.

- Train users on safe prompting and verification habits
- Review usage, incidents, and high-risk patterns regularly
- Keep guardrails aligned to new tools, agents, and business change

Your takeaway...

The strongest AI posture isn't to ban everything or allow everything. It's clear visibility, clear rules, and controls that match how people actually work.

"Prompt Injection"

When hidden or malicious instructions in prompts, files, pages, or other content try to steer an AI system away from its intended rules