# AI agents explained: how agentic AI works in production

26 October 2026 · 6 min read · blogs.iictp.org/ai-agents-explained

An AI agent is a program that uses a language model to decide what to do next, takes actions through tools, looks at the result and repeats until a goal is met. The model is the decision-maker; the rest of the system is what makes it safe and useful.

**TL;DR The short version**

- An AI agent uses a language model to choose the next step, acts through tools, checks the result and repeats until the goal is met.
- In production it needs narrow tools and caps on steps, time, spend and permissions.
- Agents compound small errors, so if the steps are known in advance, write a normal workflow and call the model only where judgement is needed.

Read online, listen to it, or change the text size: https://blogs.iictp.org/ai-agents-explained

## The loop

1. Receive a goal and the current state.
2. Ask the model to choose the next step: answer, or call a tool.
3. Run the tool (search, query, code, an API) under strict limits.
4. Feed the result back and repeat, or stop.

## What makes it work in production

- **Narrow tools.** Give the agent a few well-described actions rather than open access.
- **Limits.** Cap steps, time, spend and permissions.
- **Human checkpoints** for actions that are costly or hard to undo.
- **Logging** of every step, so a failure can be reconstructed.

## Where agents fail

Agents compound small errors: a wrong early step leads the rest astray. They can also follow instructions hidden in the content they read. Treat anything the agent reads from outside as untrusted input, and never give it more authority than the task needs.

## When not to use one

If the steps are known in advance, write a normal workflow and call the model only where judgement is needed. It is cheaper, faster and easier to test. Reach for an agent when the path genuinely depends on what is found along the way.

Related credential: AI Practitioner. Explore the exam facts and syllabus.

### Sources and further reading

1. Anthropic. “Building effective agents.” anthropic.com/engineering/building-effective-agents.

Published by IICTP, an independent technical credential body, on 26 October 2026. The current version of this post is at https://blogs.iictp.org/ai-agents-explained. More guides: blogs.iictp.org.