interweb

Generated on 15/11/2026 00:16

KNOWLEDGE BASE · PUBLIC DOCUMENT

What an AI agent is, and what it still cannot do reliably

How AI Works

CODE	VERSION	COUNTRY · LANGUAGE
RLX-AO-KB-1586	v2026.10.10	Angola · English

Interweb Angola
RLX Angola, Lda. · Group RLX Holdings
Tax ID (NIF) 500 260 5724 · Commercial registration no. 26464-25/250724

https://www.meu.interweb.ao/knowledgebase/1586

What an AI agent is, and what it still cannot do reliably

How AI Works Created on **October 10, 2026** 3 min read



An **AI agent** is a model connected to tools (read files, search, write, send messages, run commands) that works in a loop: it thinks, uses a tool, sees the result and decides the next step, until it considers the task done. A chat only answers; an agent **acts**. That is why it gets more done and why it runs more risk.

Where an agent helps and where it fails

Type of task	Typical result	Why
Short, well defined, with a clear test ("sort these files by date")	Usually goes well	Little room to drift, easy to check.
Long, with many steps	Errors add up	A wrong step early on contaminates the next ones, and the agent carries on with confidence.
Needing to know the company, the customer or the context	Weak	It only knows what it was given. The rest it invents or leaves undone.
With consequences that are hard to undo (delete, pay, send)	Dangerous without approval	It does not tell a reversible action from an irreversible one very well.
Open and vague ("grow my business")	Scattered	With no success criterion, it never knows when to stop.

The limits that do not change with the next version

1. **It can be confidently wrong.** It writes with the same confidence when it is right and when it invents. See **how to check an answer**.

2. **It can fall into a loop** and spend without producing anything. Every turn costs tokens. See **how tokens are counted**.

3 **It can be steered by a stranger's text.** A page or e-mail with hidden instructions can change what it does. See **prompt injection**.

4 **It only has the powers you give it.** That is a good limit: the less access, the less damage.

5 **It does not take responsibility.** Whoever put it to work answers for it.

How to use an agent with care

Measure	How
Minimum access	An account just for it, with permissions only for what it needs. Never your administrator account.
Human approval	It proposes, a person confirms anything that sends, pays, deletes or publishes.
Spending ceiling	A limit at the model provider, before you let it run alone.
Logs	Keep what it did and why, so you can review and roll back.
Rehearsal	First in a test environment, with nothing important in it.

If the idea is to put an agent in front of customers, first ask what happens when it gets something wrong in front of them. If the answer is "an angry customer", what you probably need is a flow with human approval, not a free-running agent.

! **An agent running on a server has access to that server.**
On an unmanaged VPS, installation, security and maintenance of the application are yours: Interweb does not install or repair third-party agents. See **how far our support goes**.

To start, choose **one small, reversible task**, give it read-only access, and compare what it does with what you would have done. Only then give it the power to write.

RECOMMENDED

Want a server to try it on, with root access? Have a look at the VPS plans.

See VPS plans →

SEE ALSO

- Hosting an AI agent: why it needs a VPS and not shared hosting →
- OpenClaw security: what it can reach, and how to limit it →
- What an MCP server is, and why people ask for one →
- VPS plans →

interweb

Interweb Angola

RLX Angola, Lda. · Group RLX Holdings

TAX ID (NIF)	500 260 5724
COMMERCIAL REGISTRATION NO.	26464-25/250724
REGISTERED OFFICE	Rua Direita do Patriota, Bairro Dangereux Belas, Luanda, Angola
CONTACT	info@interweb.ao · +244 942 011 191 · www.interweb.ao
OPENING HOURS	Monday to Friday - 08 AM to 04 PM Saturday & Sunday - Closed

Online version of this document

RLX-AO-KB-1586 · v2026.10.10 · Generated on 15/11/2026 00:16