AI BASICS // AI AGENTS

How AI agents actually think (and why it matters)

Once you understand how these systems are designed to solve problems, everything else falls into place.

> **Travis's short version:** RAG is like a highly efficient librarian: it retrieves, but it isn't thinking. An AI agent is more like a detective: it reasons, plans, takes action and adjusts as new information comes in. That difference is what moves AI beyond a rigid single-step process.

RAG: the librarian

Retrieval-augmented generation, or RAG for short, works like a highly efficient librarian. You ask a question and it quickly finds the most relevant information from a massive archive. It's fast and accurate, and if your question aligns well with the data it has access to, it gives you a really strong answer.

1. **The user delivers a prompt.**
2. **The model retrieves relevant information from a database.**
3. **The model delivers a response.**

But here's the catch: RAG isn't actually thinking, it's retrieving. It doesn't strategise, plan or adapt in real time. If a problem needs multi-step reasoning, or involves something entirely new that isn't covered in its data set, RAG hits a wall.

AI agents: the detective

This is where AI agents change the game. Unlike RAG, they don't just retrieve information: they reason, plan and take action. A librarian is great at finding information but not necessarily at interpreting or applying it. An AI agent is like a detective solving a case: gathering evidence, asking questions and adjusting as new information comes into play. Instead of just looking for the best answer, an agent can ask *how should I solve this?* and then figure it out step by step.

RAG (the librarian)	AI agent (the detective)
Retrieves the most relevant information from an archive. Fast and accurate when the question matches the data. Doesn't strategise, plan or adapt in real time, and hits a wall on multi-step reasoning or anything new.	Reasons, plans and takes action. Gathers evidence, asks questions and adjusts as new information arrives, step by step, until the case is solved.

Two levels of thinking

AI agents operate on two different levels:

Level	How it works
Fast, reactive thinking	Like muscle memory. If an agent has seen a similar problem before, it can pull from past experience and solve it quickly.
Strategic, step-by-step thinking	For a novel challenge, the agent slows down, breaks the problem apart and formulates a plan before acting. It uses a reason-and-act loop to keep improving, and, like a detective, it doesn't stop until the case is solved, or until it hits a dead end and has to rethink its approach.

The compound systems approach

A major shift is happening towards compound systems, where AI agents don't rely on a single model but integrate multiple models and tools to solve problems step by step. Instead of being restricted to a single source, an agent can:

1. **Think and plan.** Before jumping to an answer, it strategises how to tackle the problem.
2. **Take action.** It uses the best tool for the job, whether that's searching, running calculations, writing code or even calling on another model.
3. **Generate insights.** It analyses results and refines its understanding, rather than just retrieving information.
4. **Adapt and iterate.** If the first approach isn't effective, it reassesses, tweaks its plan and tries again, learning from each attempt.

The bottom line

The agentic approach is what lets AI move beyond a rigid single-step process. Agents are becoming active problem solvers: they strategise, take action and redefine their approach on the fly, which lets them break down complex challenges, automate multi-step workflows and optimise decisions. **Next:** how agents are being put to work in your industry.

Want to talk this through for your business?

Book a free 15-minute call. We'll look at your situation and give you a straight answer, with no jargon and no sales pitch.

Let's Chat

Opens the request form on our website. Free, no obligation.