AI AGENTS

A Goldilocks Introduction to AI Agents: Opportunities, Challenges, and Everyday Impact

Michael Tefula
06 Dec 2024 — 7 min read

AI agents have quickly evolved from obscure technical experiments to mainstream buzz. In the last 18 months, new automation capabilities have captured the imagination of entrepreneurs and established organizations alike.

While it's possible that we're in another AI bubble, the early adopters of AI agents are reporting impressive results: million-dollar cost savings, automations that complete tasks in half the time, and AI agents that do the work of hundreds of humans.

This innovation presents a fundamental shift in how to automate complex work. However, there are stark limitations, challenges, and lots of open questions.

Google Search Data

Introduction

In this post, I'll share a "goldilocks" introduction to AI agents that provides a grounded take on the technology. Unlike most content on the topic, this post avoids technical jargon without being too simplistic.

I've also reviewed over 100 startups building products in this area. With that foundation, I'll share an accessible definition of AI agents, real use-case examples, and key opportunities and questions to consider.

If you find the post useful, please share it with friends and colleagues. Feel free also to reach out if you have any questions or comments.

What are AI agents?

AI agents are software systems that use AI to plan, coordinate, and execute complex tasks using tools, memory, and other AI systems to achieve their objectives.

These systems operate in ways similar to humans. For example, if you set an ambitious goal, the first step is to think, plan, and break the objective into smaller, concrete steps.

You then pick the right tools to work on each task, observe the results, and update your approach and progress until you achieve your objective.

An overview of an agentic system (Lilian Weng, 2023)

That's how AI agents work. The principles are simple: plan, execute, review, repeat. However, simple principles can achieve extraordinary results if the systems coordinating the process have access to intelligence, tools, and memory. Recent breakthroughs in large language models (LLMs) and generative AI (gen AI) have made this possible.

A smart LLM equipped with additional software tooling can autonomously plan, critique, delegate, execute and coordinate tasks until it completes an objective. What does this look like practically?

Let's consider a simple code example that anyone can follow.

A Simple Agentic System

A friend recently asked me whether an app could find you the best-priced airline tickets for the best weather destination. This is possible with traditional code. However, AI agents can provide a more engaging approach that uses natural language.

Here's a quick example of an agentic workflow that provides travel advice. If you don't know how to code, don't worry. I'll break the steps down into four simple processes.

1. **Get user preferences:** Users can tell the AI their weather preferences and departure location using natural language. LLMs act as the brain that takes your words and converts them into useful computer information. The AI will reflect on what you say and figure out what tools it needs to fulfill your request.
2. **Find destinations:** Once the AI understands what the user wants, it can use a tool called “find_destinations”. This tool helps the AI match your weather preferences against the weather data of over 200 cities for the last 12 months. Here, the AI is augmenting its own capabilities with fresh data.
3. **Find flights:** When the AI has a list of locations with the perfect weather for the user, it can then leverage another tool to get flight data and ticket price information.
4. **Generate recommendations:** Armed with the data from steps 2 and 3, the AI now uses its language skills to summarise all the data into a recommendation.

Here's a sample result when I run the full code:

> *Cairo, Egypt*

> *Cairo is a fascinating city with a rich history and vibrant culture. With the weather boasting an impressive 3,917.71 hours of sunshine annually and minimal rainfall at just 27.3 mm, it's a great destination for those who love sunny days and exploring ancient wonders like the Pyramids of Giza and the Sphinx.*
>
> *Flight Options:*
>
> *Wizz Air offers a flight at £207. It departs on December 5, 2024, at 14:10 and arrives at 21:15. This is the shortest flight duration available, making it a convenient choice.*
>
> *Vueling has a slightly cheaper option at £206, but with a longer duration. It leaves on December 7, 2024, at 13:10 and arrives at 22:25."*

Advanced Agentic Systems

Google's Presentation on AI Agents

Sophisticated agentic systems are now far more capable thanks to how powerful LLMs have become. These systems are already being used in both large and small organizations. The areas gaining the most adoption so far include:

- **Software development:** AI agents that build and maintain software (e.g. Devin, Cursor, Replit, or even Microsofts Github Copilot, which now has 1.8m subscribers).

- **Customer service:** AI agents that handle customer queries (e.g. Klarna's AI agents do the equivalent work of 700 employees and are forecast to save $40m of costs in 2024.)
- **Sales and marketing:** AI agents that automate sales prospecting and marketing (e.g. KFC and Taco Bell's owners are experimenting with gen AI and have generated double-digit increases in consumer engagement).

Areas of Opportunity

I reviewed 130+ startups building AI agents, initially focusing on a batch of businesses that go through the world-renown Silicon Valley startup programme Y Combinator. This work revealed a diversity of use-cases with lots more to come.

You can view an updated interactive version of this market map here.

In the applications category I found systems that can assist doctors (e.g. OpenClinic), train robots (e.g. innate), act as your second brain (e.g. Khoj), or even work as your personal interior designer (e.g. Rastro).

Although some use-cases are speculative, others are promising. One example is HappyRobot. It uses AI agents to automate phone calls and communications for logistics businesses. Their 50+ customers are already seeing tangible benefits. Call times have been cut by 50% while operational costs have come down by a third.

On the tooling side, I found platforms that can help you build AI agents without writing a line of code (e.g. Gumloop), systems that enable agents to take payments securely over the phone (e.g. Protegee), and platforms that use “red-team” AI agents to automatically stress test agentic systems to ensure their security and alignment with human objectives.

Current Limitations

There's lots of hype around AI agents today so it's worth grounding our expectations by taking into account current limitations. Some of these issues will be resolved in the near future but others may elude us for longer than we expect. Examples include:

- **Technical limitations:** The brains of AI agents (LLMs) are prone to reliability issues. They can randomly make things up (hallucinations) and they struggle to plan and reason over long horizon objectives. This risks compound errors when chaining multiple tasks together. A 10-step process that's 90% accurate at each juncture will be just 35% reliable at the end (90% ^ 10).
- **Operational challenges:** AI agents that can interact with other software systems, manage sensitive information, and even make autonomous decisions– such as executing or accepting payments–present significant integration challenges, along with data privacy and security concerns. However, the connective tissue and safeguards required for such systems are still underdeveloped.
- **Human & organizational barriers:** It will be a while before society can trust AI agents enough to adopt them en-mass. Aside from the reliability and safety concerns, there's also the issue of AI eliminating jobs and disrupting our way of work. Full automation might be possible in some areas but it won't necessarily be desirable.

Open Questions

The opportunities that AI agents present are significant and will no doubt transform economies. However, as these systems become more intelligent and pervasive we will have to grapple with several open-ended questions. Here are a few I've pondered.

Technical questions

- Is the future all about specialist AI agents or is there a path to a high-performing generalist AI Agentic system?
- Are we going to continue using LLMs as the brains of agentic systems despite their probabilistic nature or do we need a more deterministic planning system?
- What benchmarks should we use to assess performance? Equivalent human performance or something better?

Human and workforce impact

- How much will and should humans stay in the loop? Where is it important, where is it less critical?
- How will AI agents impact employment and the global workforce?

Business models and commercial issues

- How should AI agentic products be priced? Per task? Per hour? Per incremental unit of value created?
- What new creative things can we do with AI agents that couldn't be done before? For example I can code up an agentic system that can synthesize thousands of product reviews – something a human couldn't do. What else is possible at scale and can generate meaningful positive impact?

Regulations and risks management

- How should agentic systems be regulated? What about privacy and security issues?
- Who's liable when an agentic system goes wrong and causes real damage?

There are more questions to explore than I can fit into this blog post. So for now, I would encourage you to try out the technology and think through some questions of your own. How can AI agents help you do better work and play? What's hype, what's real today, and what does the future hold? I look forward to learning more.

READ MORE

Open Problems (& Opportunities) for AI: Summary Notes from the Conference

The Software Engineer is Dead, Long Live the Software Engineer!

An undeniable killer use case for large...

Day 100 of 100 Days of AI 🎉

AI agents have been getting a lot of...

03 Jul 2024

Day 99 of 100 Days of AI

I switched one of my AI agents projects...

02 Jul 2024

Last week I attended the Algorithmic...

01 Nov 2024

25 Sep 2024

Michael Tefula