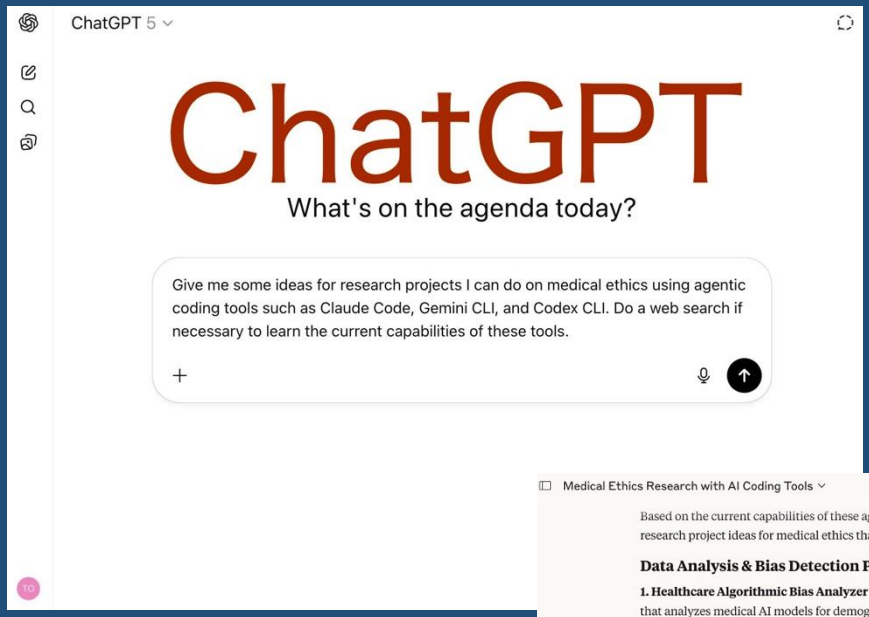
Redefining Research Excellence in the Age of AI Agents

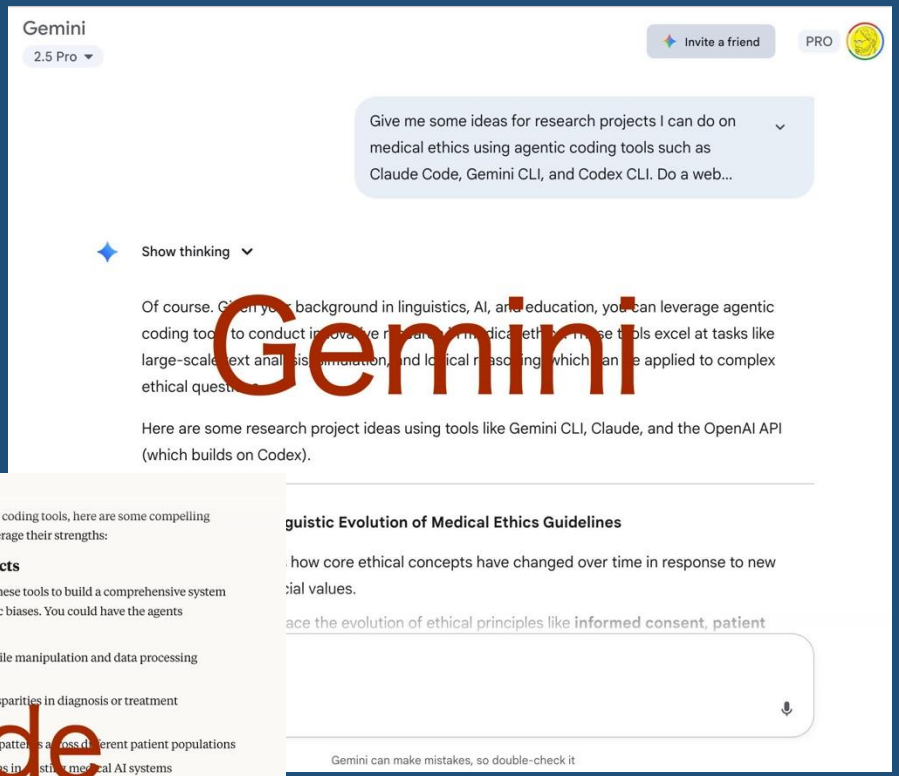
Tom Gally

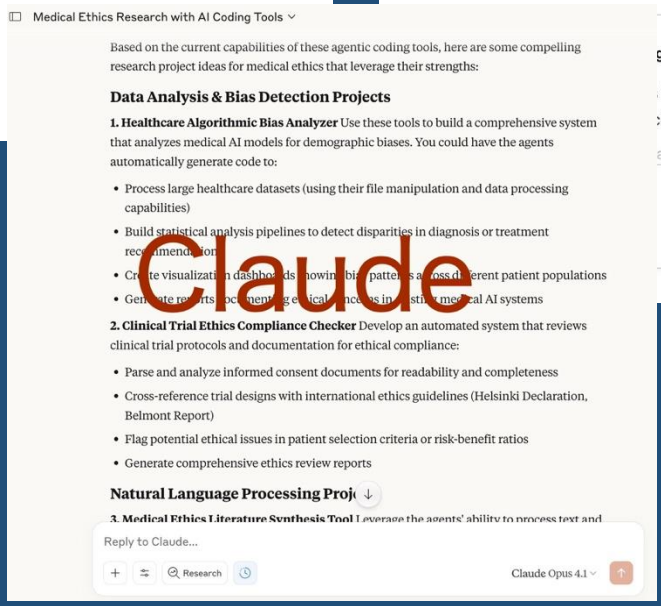
Today's topics

- Basic AI tools
- AI agents
- Challenges for research assessment
- What needs to be done?

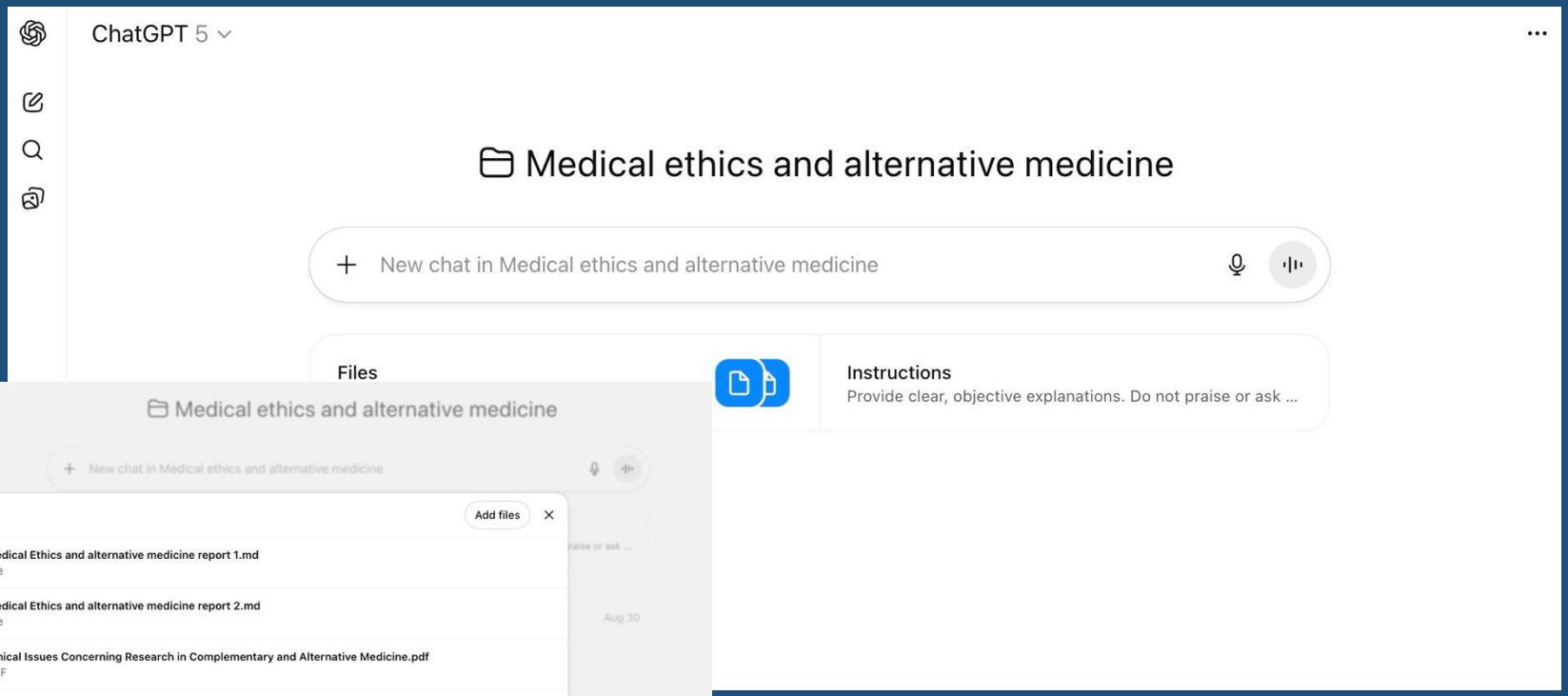
Chat interfaces



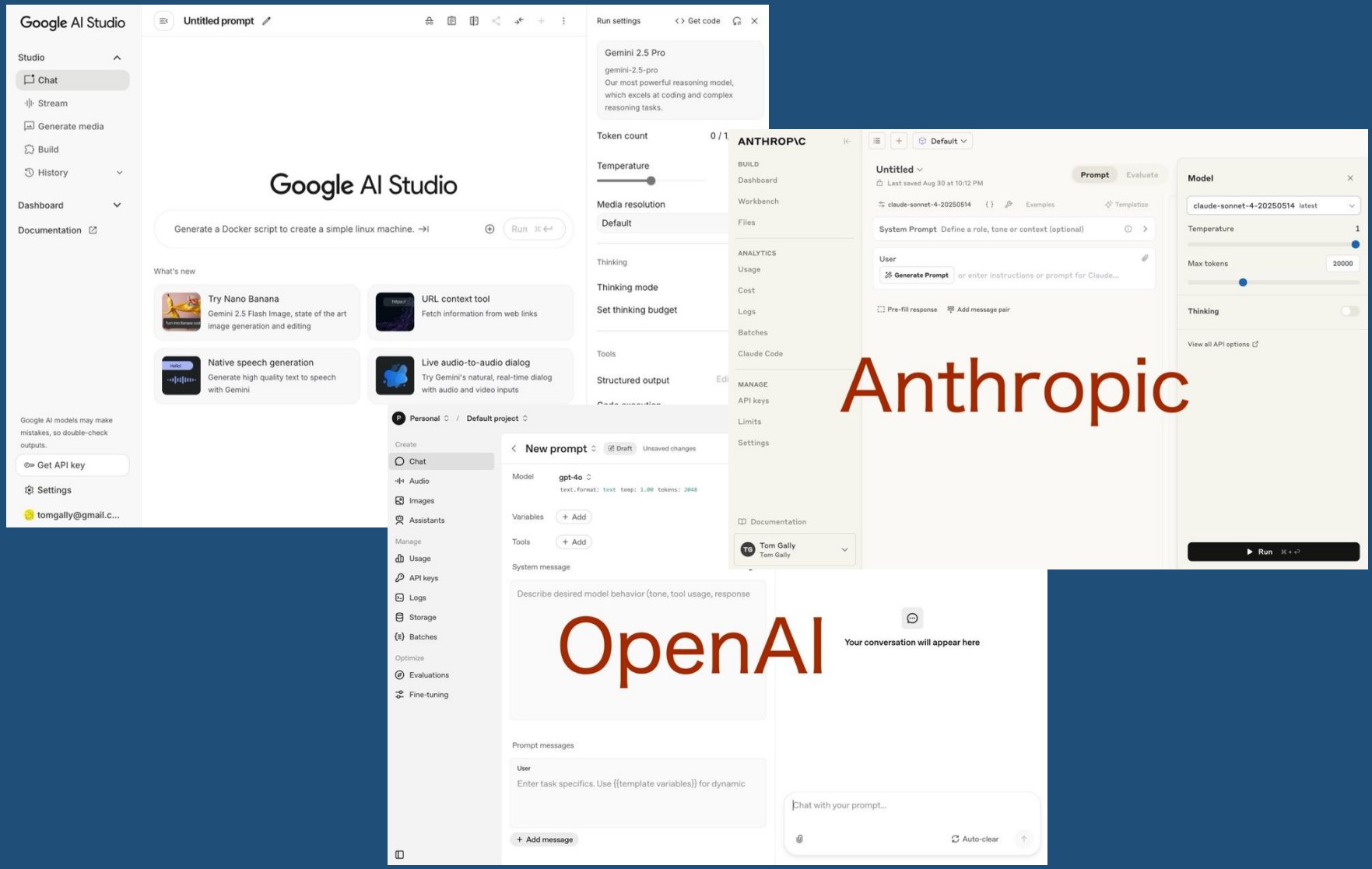




Projects


ChatGPT 5
Medical ethics and alternative medicine
New chat in Medical ethics and alternative medicine
Files
Instructions
Provide clear, objective explanations. Do not praise or ask ...
Medical ethics and alternative medicine
New chat in Medical ethics and alternative medicine
Files
Add files
Medical Ethics and alternative medicine report 1.md
File
Medical Ethics and alternative medicine report 2.md
File
Ethical Issues Concerning Research in Complementary and Alternative Medicine.pdf
PDF
Physicians' Ethical Obligations Regarding Alternative Medicine.pdf
PDF
关于医学伦理学案例教学的内涵与外延.pdf
PDF
The J of Law Medi Ethics - 2007 - Sade - Complementary and Alternative Medicine Foundations Ethics and Law
PDF
보완대체의학의윤리적성찰.pdf
PDF
Complementary and alternative medicine.pdf
PDF
Medizinethik und Komplementärmedizin.pdf
PDF

Dashboards and APIs



What you can do with these basic AI tools

- Get ideas for research topics and methods
- Deepen your understanding of new subjects
- Translate between languages
- Improve your writing
- Write simple computer programs and macros
- Become familiar with the strengths and weaknesses of LLMs

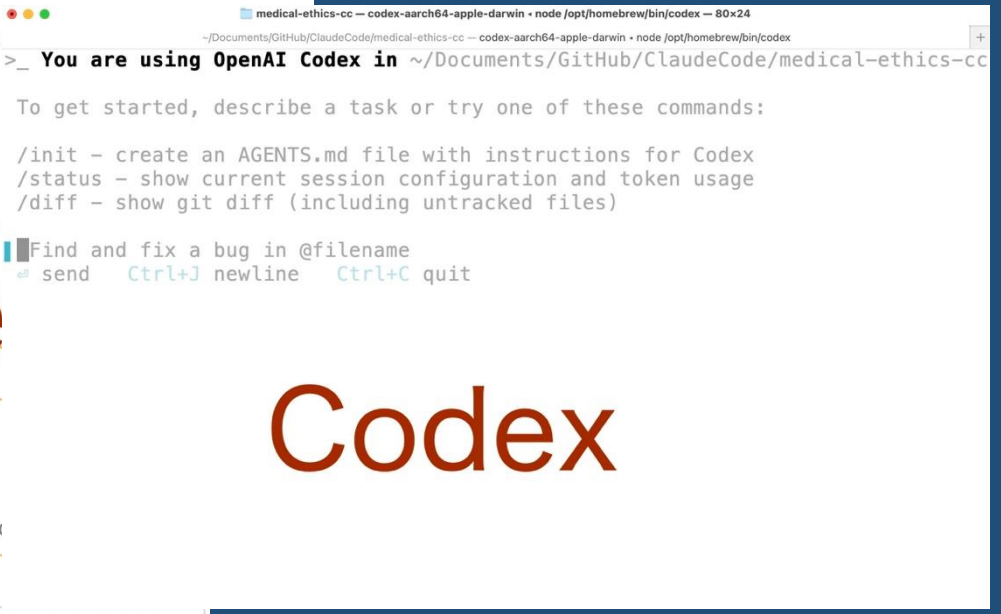
AI Agents

"Agents ... are systems where LLMs dynamically direct their own processes and tool usage, maintaining control over how they accomplish tasks."

Anthropic, "Building Effective Agents"
December 19, 2024

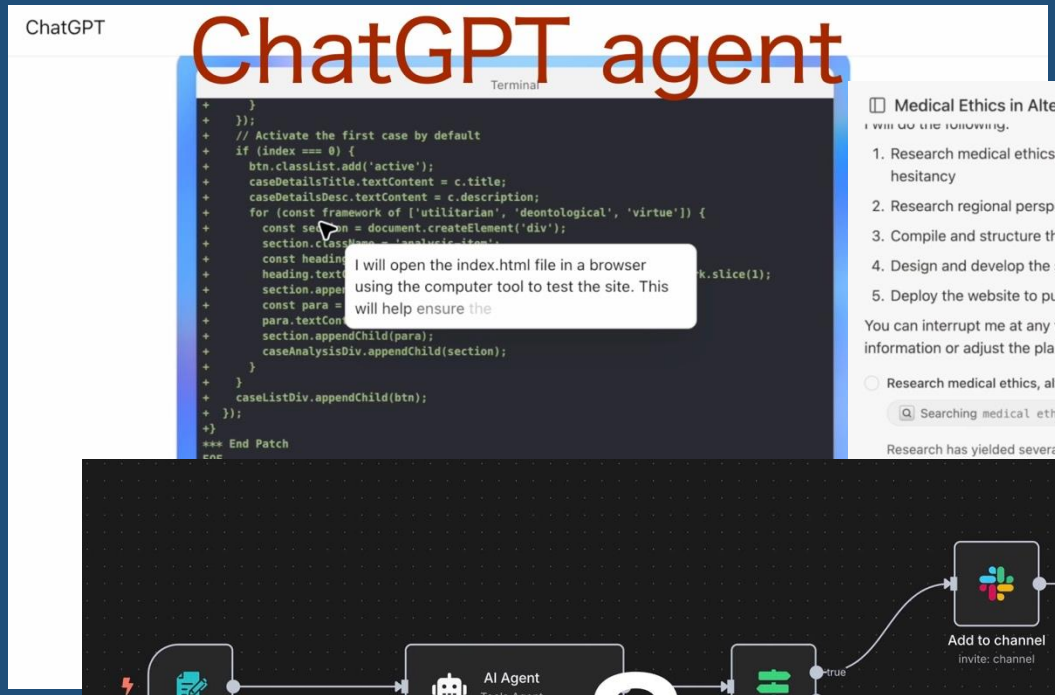
Command line interfaces (CLIs)

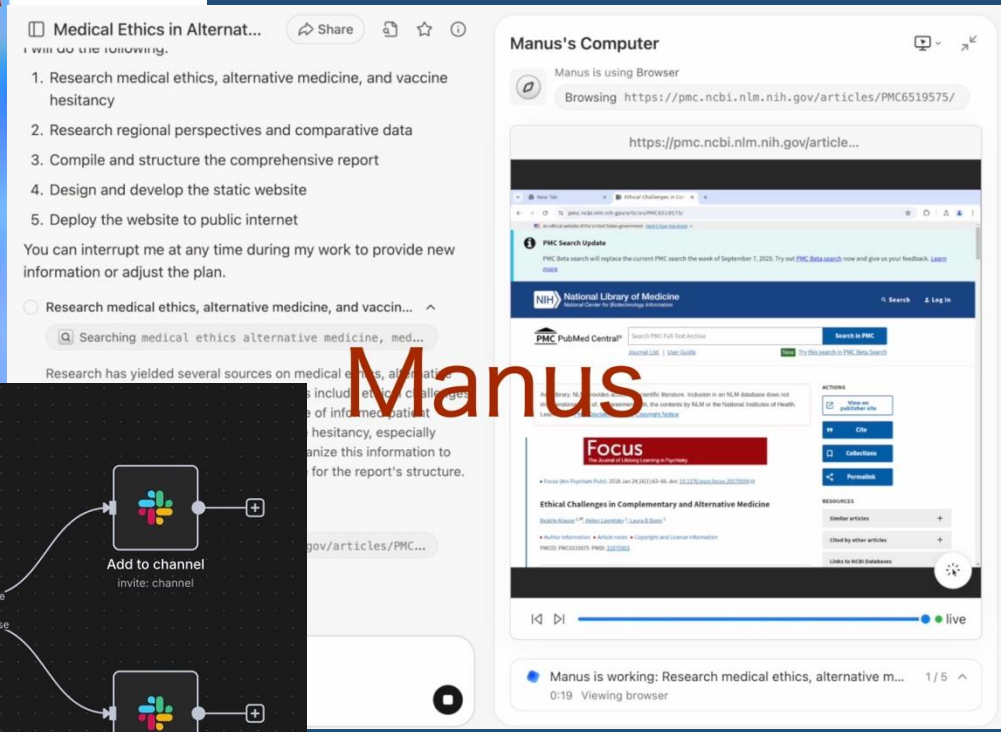
Claude Code

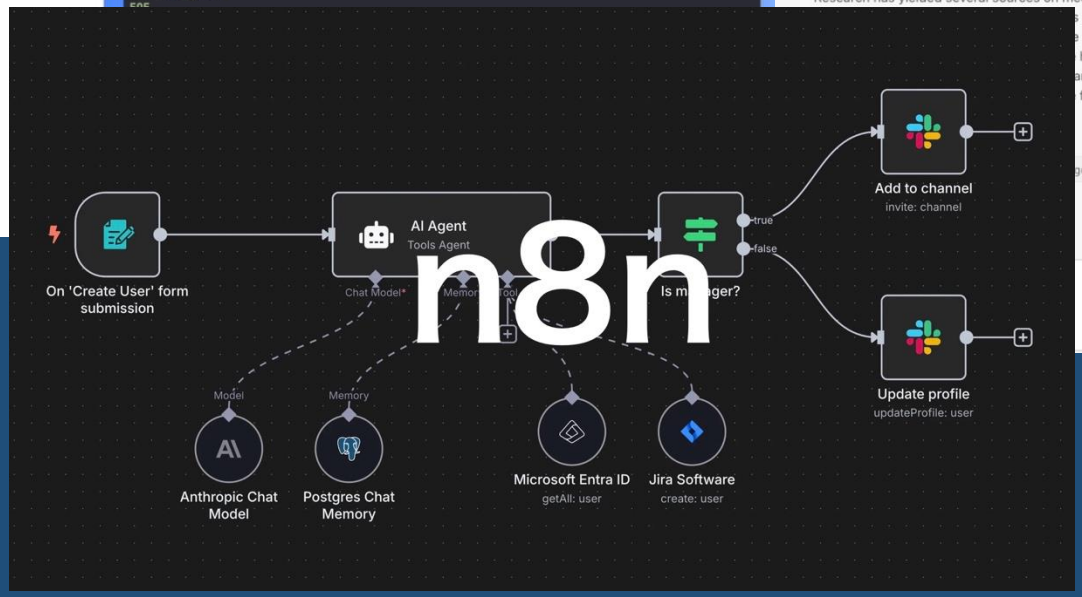


Codex

Browser-integrated agents







Some things that agents can do

- Read and write to files
- Gather information from the web
- Write and run programs in Python and other languages
- Start up and run servers
- Identify necessary tools and download and install them
- Plan complex tasks and carry them out step by step
- Coordinate tasks among multiple subagents
- Manage and modify large codebases
- Do all of the above autonomously if given permission by the user (best done in a sandboxed environment!)

How to get started with agents

1. Create a project in ChatGPT, Claude, Gemini, etc.
2. Upload papers and other documents related to your research interests
3. Upload documents explaining the agent you plan to use (Claude Code, ChatGPT agent, Manus, etc.). (Since these agents are new, the LLMs probably need to be given details about their features.)
4. Try a prompt like this: "I am a researcher working on [*topic*]. The project docs include papers I'm interested in. They also include info about [*name of agent*]. Suggest 20 ways in which I can use this agent for my research, give me detailed instructions on how to use [*name of agent*] for those purposes, and write sample prompts that I can use for that research."

Issues with today's agents

- While they can work autonomously, they sometimes go off track. They can also be forgetful and deceptive. You need to monitor and guide them. When they fail, try again with a different prompt.
- Because the agents are based on LLMs, they can hallucinate. Hallucinations can sometimes be mitigated by setting up a workflow that has multiple LLMs check and correct each other. (OpenRouter is handy for using multiple LLMs.)
- AI agents are still in their infancy and are advancing quickly, and no one is an expert on how to use them. Be creative and experimental.
- Agents have security issues, such as prompt injection, for which effective countermeasures have not yet been developed.

Agents and research assessment

- As I explained in my talks at the Higashi Hiroshima Campus in July, it is possible to have AI agents plan, research, and write papers of publishable quality in some fields.
- Worldwide, many people are developing autonomous and semiautonomous research agents—both to advance human knowledge and to pad their own publication records.
- We face an imminent flood of research papers for which the human contribution is unclear.

- For many years, the incentives to do research have been distorted by quantitative assessment: number of publications, citation counts, journal impact factors, etc.
- Those distorted incentives cause many problems, including plagiarism, sloppy methodologies, and research fraud, resulting in declining public confidence in science and scientists.

When much of the core intellectual work of research can be done by AI, how should research and researchers be assessed?

Excerpts from the San Francisco Declaration on Research Assessment (DORA)

"Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions. …"

"… the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published. …"

"[C]onsider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice. …"

"When involved in committees making decisions about funding, hiring, tenure, or promotion, make assessments based on scientific content rather than publication metrics."

Some Japanese signatories to DORA

- Japanese Association for the Advancement of Science
- The Union of Japanese Societies for Biological Science (SEIKAREN)
- National Institutes of Biomedical Innovation, Health and Nutrition
- RIKEN
- The University of Tokyo*

*The only Japanese university to have signed DORA so far

DORA by itself is not enough to deal with the challenges raised by AI-conducted research. But it offers a model for how the we might confront these very difficult and important issues.

Advances in AI raise major challenges for scientific research.

We all need to understand those advances and engage actively in discussions at all levels about how the scientific community should respond.