WHAT ARE WE UP AGAINST IN ACTIVE INFERENCE?

Chris Mathys

International Workshop on Active Inference

Montréal, 15 October 2025

INTERACTING MINDS CENTRE (IMC)
AARHUS UNIVERSITY

OUTLINE

- Do we really have to do philosophy? Can't we just get on with the stuff that actually matters?
- But it matters what we do and why we do it. So it matters to address a number of recurring objections to what we do when we use active inference as an approach to constructing artificially intelligent agents.
- I will start from some of the superficial objections we encounter and work from there to uncover the underlying assumptions and convictions that drive these objections.
- Borrowing insights from theoretical economics, which also had to clarify its epistemological foundations, I will argue that there are two main problems we are up against
- One is grounded in misconceptions that are hard, but possible, to correct
- The other is grounded in the nature of active inference, and we face it regardless of misconceptions

SUPERFICIAL OBJECTIONS

- Cognitive science (and thereby active inference) has been **made redundant by generative AI** (e.g., the generative pre-trained transformers of ChatGPT fame). Remaining problems will be solved automatically by scaling.

- We don't need anything as slow-moving as active inference since we've got **machine learning**

- We don't need anything as complicated as active inference since we've **got reinforcement learning.**

- Active inference is **not falsifiable** and therefore not scientific.

- These have all been addressed, but they keep coming, and there is more behind this than obtuseness in the objectors

WE'VE GOT MACHINE LEARNING!

Tal Yarkoni • 2nd
Research Scientist at Midjourney
1w •

But the crux of the talk was really about a question I viewed as kind of existential for the field. ML/AI researchers, I pointed out, were now routinely doing things that we psychologists had long thought of as science fiction. Things that we'd been telling funding agencies we might solve in 50 years, if only they gave us enough money to *really* test our theories. What would we do, I asked, if ML folks solved those problems in 5 or 10 years, no psychology required?

WE'VE GOT MACHINE LEARNING!

Tal Yarkoni • 2nd
Research Scientist at Midjourney
1w •

Still, I think the 6 or 7 years since I gave that talk have affirmed the general shape of the premonition. I no longer sit on grant panels, so I don't know how psychologists are pitching their work to the funding agencies these days. I do know that many of the huge problems I used to read about in Specific Aims pages have been, or are being, solved—by ML practitioners.

WE'VE GOT REINFORCEMENT LEARNING!

The Bitter Lesson

Rich Sutton

March 13, 2019

"The bitter lesson is based on the historical observations that 1) AI researchers have often tried to build knowledge into their agents, 2) this always helps in the short term, and is personally satisfying to the researcher, but 3) in the long run it plateaus and even inhibits further progress, and 4) breakthrough progress eventually arrives by an opposing approach based on scaling computation by search and learning. The eventual success is tinged with bitterness, and often incompletely digested, because it is success over a favored, human-centric approach."

http://www.incompleteideas.net/IncIdeas/BitterLesson.html

WE'VE GOT REINFORCEMENT LEARNING!

The Bitter Lesson

Rich Sutton

March 13, 2019

"... the actual contents of minds are tremendously, irredeemably complex; we should stop trying to find simple ways to think about the contents of minds, such as simple ways to think about space, objects, multiple agents, or symmetries. All these are part of the arbitrary, intrinsically-complex, outside world. They are not what should be built in, as their complexity is endless; instead we should build in only the meta-methods that can find and capture this arbitrary complexity. Essential to these methods is that they can find good approximations, but the search for them should be by our methods, not by us. We want AI agents that can discover like we can, not which contain what we have discovered. Building in our discoveries only makes it harder to see how the discovering process can be done."

http://www.incompleteideas.net/IncIdeas/BitterLesson.html

BUT...

Gary Marcus @GaryMarcus · Sep 26

Astonishing. It's been a long, hard, unpleasant road.

But one by one, every major figure in AI has come around to the critique of LLMs that I began giving here in 2019.

@ylecun was first.

Sir @demishassabis sees it now, too.

Turing Award winnner @RichardSSutton, famous for

Show more

> **Gary Marcus** @GaryMarcus · Sep 26
>
> What has this world come to?!
>
> Mr Bitter Lesson, Richard Sutton, now sounds exactly like ... me.
> x.com/iamaniku/statu...

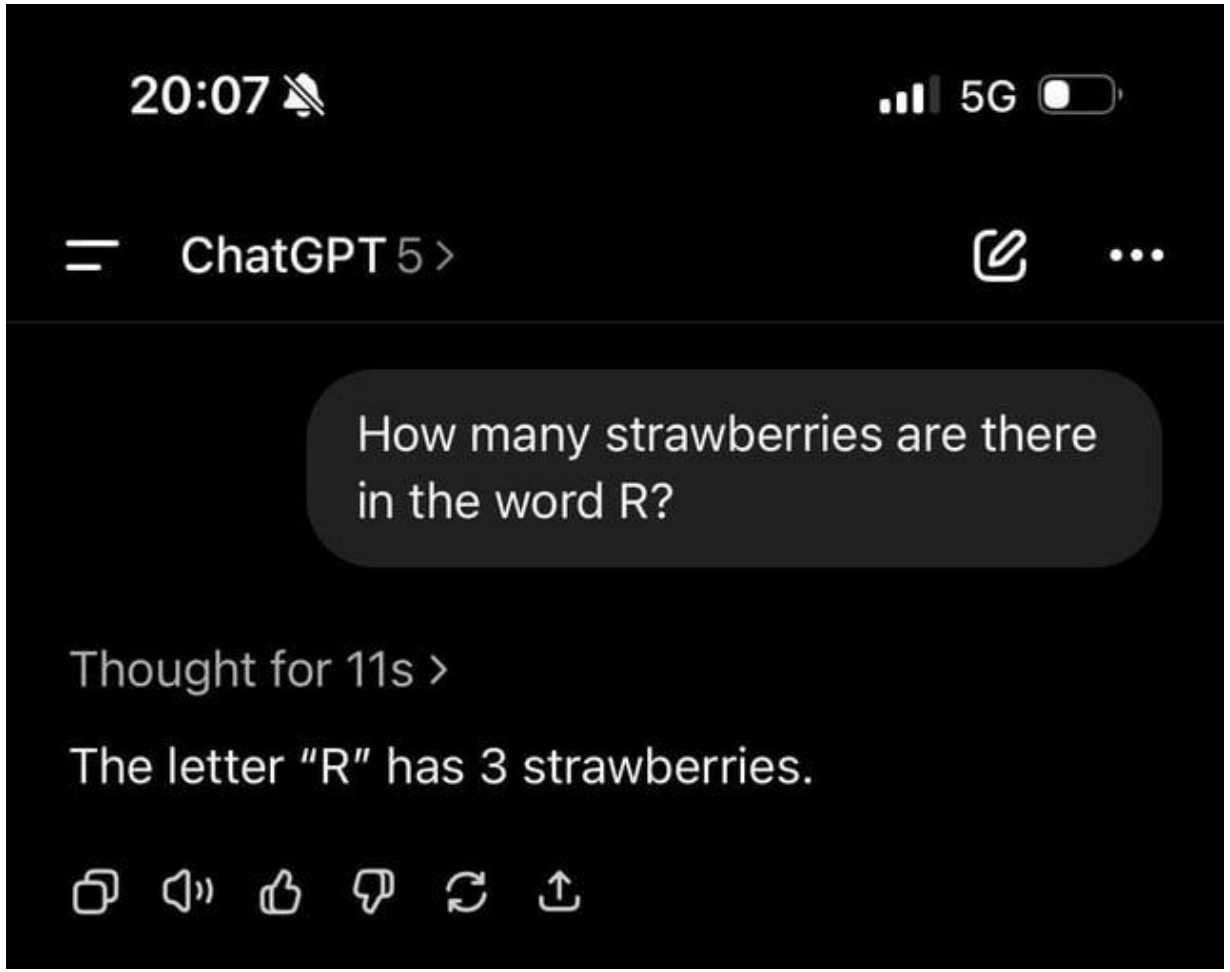
69 121 908 235K

Richard Sutton @RichardSSutton · Sep 27

You were never alone, Gary, though you were the first to bite the bullet, to fight the good fight, and to make the argument well, again and again, for the limitations of LLMs. I salute you for this good service!

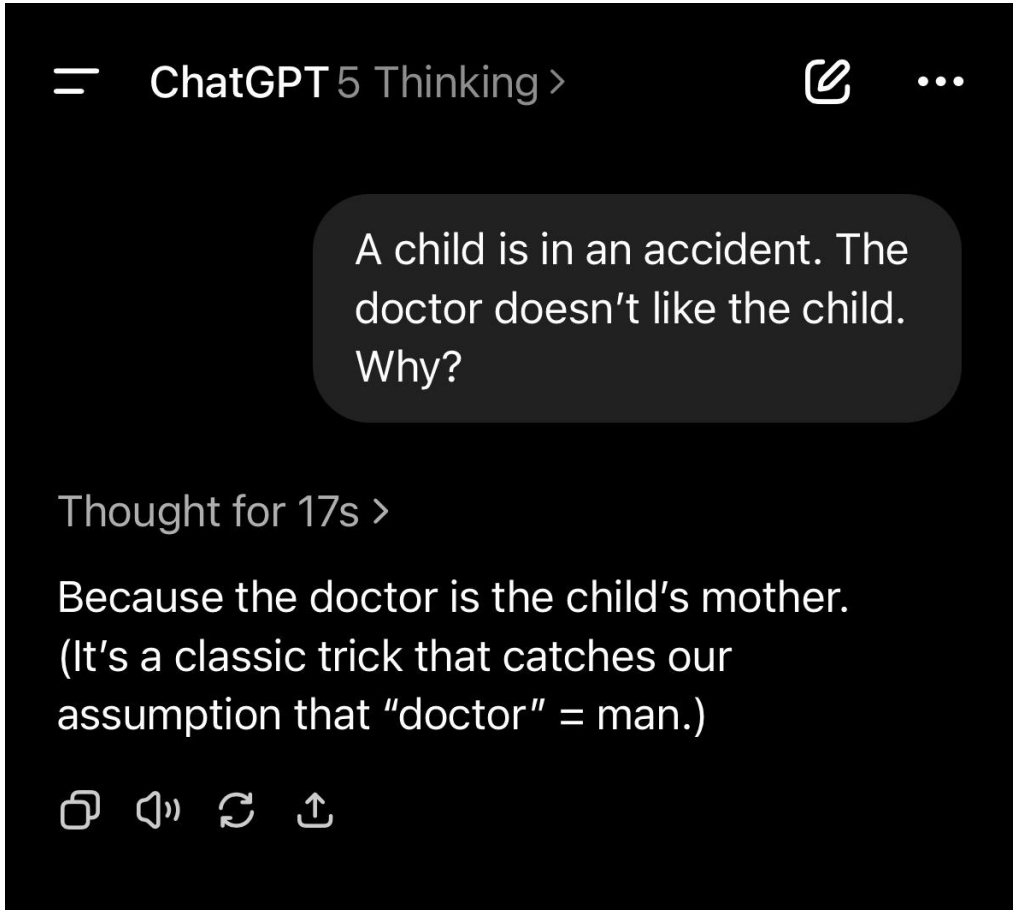
40 157 1.2K 661K

WHAT LIMITATIONS OF GENERATIVE AI?



https://www.reddit.com/r/ChatGPT/comments/1naxubm/dont_worry_our_jobs_are_safe/

WHAT LIMITATIONS OF GENERATIVE AI?



https://x.com/Wasgo/status/1954375973044101175

WHAT LIMITATIONS OF GENERATIVE AI?

https://www.reddit.com/r/ChatGPT/comments/1czchmq/why_cant_chatgpt_play_chess/

WHAT LIMITATIONS OF GENERATIVE AI?

The Illusion of Thinking:
Understanding the Strengths and Limitations of Reasoning Models
via the Lens of Problem Complexity

Parshin Shojaee*† Iman Mirzadeh* Keivan Alizadeh
Maxwell Horton Samy Bengio Mehrdad Farajtabar

Apple

EPISTEMOLOGY AND HUMAN ACTION

- The failures of generative AI, and the implied failure of the critiques of active inference and cognitive science in general that are based on the unreasonable expectations underlying these failures point to **deep-seated misunderstandings about the nature of cognition**.

- How do the agents we describe and construct know anything, and what does knowing something mean? These are questions of **epistemology.**

- Another field facing precisely the same questions is **economics,** which describes **humans acting in order to achieve goals.**

- Accordingly, the important 20th-century economist Ludwig von Mises' magnum opus is called 'Human Action' (1949), and he has a few things to say that turn out to be relevant to the questions above.

EPISTEMOLOGY

- Three kinds of epistemology: rationalism, empiricism, and historicism (Hoppe, 1993)
- **Rationalism** (roughly): some of what we know, we know by experience, but there are things we know a priori, and these are fundamental to our experiential acquisition of knowledge
- **Empiricism:** "Nothing is in the intellect that has not previously been in the senses" (Locke)
- (Leibniz's rationalist rejoinder: "... except the intellect itself")
- **Historicism:** everything is like a literary text, there are no constant relations

Hoppe, H.-H. (1993). *On Praxeology and the Praxeological Foundation of Epistemology* in Meaning of Ludwig von Mises: Contributions in Economics, Epistemology, Sociology, and Political Philosophy, Jeffrey M. Herbener ed. (Auburn, Ala.: Ludwig von Mises Institute).

EPISTEMOLOGY

- Both empiricism and historicism are self-contradictory.
- The contradiction in empiricism (roughly): the claim that everything is based on experience is not itself based on experience.
- The contradiction in historicism (roughly): if there are no constant relations, nothing of any substance can be said at all, not even that there are no constant relations.
- The consequence of empiricism is a **skeptical** world-view, where anything we know could be invalidated on the next observation, everything is always provisional.
- The consequence of historicism is a **relativist** world-view, where 'my' truth is as good as 'your' truth.
- It is easy to see how skepticism relativism can lead to the same – relativist – place (and a worship of power since in a relativist world the biggest bully ends up imposing 'his' truth)

RATIONALISM AND ACTIVE INFERENCE

- So we are stuck with **rationalism**
- This is to say we are stuck with **active inference**. Why?
- The epistemology of human action rests on **two axioms (Hoppe, 1993)**
- First, the **axiom of action: humans display intentional behaviour. (**Mises, 1949)
- Cannot be denied since denying it would itself be intentional behaviour.
- Hoppe: **“But is this not just trivial?”**
- No, since in this psychologically trivial axiom there are **insights implied which are not themselves psychologically self-evident** (e.g. hierarchy of preferences, temporality of decisions, etc.)

RATIONALISM AND ACTIVE INFERENCE

- "All of these categories which we know to be the very heart of economics—values, ends, means, choice, preference, cost, profit and loss—are implied in the axiom of action." (Hoppe, 1993)

- Second, the **axiom of argumentation: humans are capable of argumentation and hence know the meaning of truth and validity.**

- Cannot be denied since denying it would itself be arguing about a truth claim.

- The combination of these two axioms puts **action** front and center in epistemology: **being an actor (agent) and knowing something depend on each other.**

- The combination of the axioms also solves a long-standing problem of rationalism: **how can we know that what we know a priori** (independent of experience) **corresponds to reality?**

- Answer: **all knowledge is the knowledge of an agent,** validated by (and enabling) action. It is knowledge that cannot simply not be thought to be otherwise, but also and more importantly **knowledge that cannot be undone**. There is nothing you can **do** to avoid having four items after putting two and two of them together.

REVISITING THE OBJECTIONS

- The above objections to active inference are rooted in the **fallacy of empiricism.**

- **"The LRM (language reasoning model) will have to acquire (at least) human capabilities if we only scale enough"** – i.e., if we increase its experience enough. We now know from first principles that **this cannot work.** Some of our knowledge is a priori, and that **knowledge is bound up with action.**

- "But are these examples of yours recent? **Hasn't this already been fixed? When I talk to ChatGPT it seems much, much smarter than in your examples.** I'm sure these are just minor glitches that'll be ironed out soon." The ELIZA effect: passing the Turing test (i.e., being a capable bullshitter) is much easier than being a capable actor. Notably **the Tower of Hanoi is an action problem. Playing chess is an action problem.**

REVISITING THE OBJECTIONS

- **An 'agent' cannot learn to act simply by observing actions.** It needs an a priori understanding of what it means to be an actor.

- **An LRM cannot learn to reason simply by observing others making statements.** It needs an a priori understanding of what it means to make a truth claim.

- In sum, learning by example (and scaling that up) cannot work in an agent which (like all of empiricism) has **no notion of the impossible** (cf. Chomsky).

- If everything we know can always be invalidated by the next example, then **anything is possible**. The result of empiricism is **not humility**, as the empiricists would have it, **but hubris** (e.g., violation of the rules of physics in AI-generated videos, violations of the rules of chess, etc.)

WHAT WE'RE UP AGAINST

- Given the way knowledge and action depend on each other, active inference (in a broad sense) is without doubt the right approach to artificial intelligence.

- **However, we are up against two big things:**

- **First, an intellectual environment steeped in empiricist prejudice and the attendant hubris.** This can be addressed by arguing from first principles based on knowing where active inference stands epistemologically.

- **Second, the limits of our introspection.** We cannot simply go and equip an artificially intelligent agent with the a priori knowledge that we as naturally intelligent agents have. This a priori knowledge is only partly accessible to introspection, and uncovering it will require hard work.

Thanks

AARHUS
UNIVERSITY