

AI2Reason

Build and characterize artificial reasoning system that can reproduce human-level reasoning abilities

By Zory Zhang

Speaker notes: Version: elevator pitch (30s), talk with professors (10min), talk with interested senior (20min), talk with potential collaborators (40min). Audience can ask for more elaboration on any part of the talk after the standard presentation.

Outline

Goal: introduce and promote AI2Reason as my / a promising AI research direction.



1. What's AI2Reason
2. Why important at this moment
3. Why is it hard but promising now
4. My Next step

1 What's AI2Reason

Outline recap:

1. **What's AI2Reason**
 - A. Goal
 - A. Some key features
 - B. What aspects of intelligent system are covered?
 - C. What aspects of intelligent system are not covered?
2. Why important at this moment
3. Why is it hard but promising now
4. My Next step

A. Goal

- **Reproduce human-level reasoning abilities**
-  just stronger computational model
-  **characterize** how an artificial reasoning system can be / needs to be like

B. Some key features

Speaker notes: To exemplify this idea, I will cover three examples of reasoning abilities at different level ...

From **solving math word problems**

- formulate / translate (not covered) → planning → automated theorem proving

Speaker notes: just like how mathematicians sometimes seriously treat the problem. They first formulating it into abstract math question first (translate into formal language, which is not covered), planning on how to solve it (planning), and then solving each subgoal one by one (automated theorem prover)

To serve as a **scientific enquiry assistant**

Speaker notes: given a phenomenon of interest, give hypothesis with explanatory power, and conduct thought experiment / propose real world experiment to confirm/develop/deny it

- ~~Meta AI Galactica: "store, combine and reason about scientific knowledge"~~
- observer
- hypothesis generation
- thought experiment / real world experiment
- reasoning on hypothesis as explanation

To **develop new theory**

- Scientific concept / diagram change.
- E.g. weight of object → universal gravity

C. What aspects of intelligent system are covered?

- **Deductive, inductive, abductive reasoning.**
- **Categorization and conceptualization.**
- **Planning.**
- **Causality.**
- **Explanation seeking.**
- (All of them are exemplified in doing math)

D. What aspects of intelligent system are not covered?

Speaker notes: (We are not as ambitious as you might think)

- **Perception / visual reasoning / embodied reasoning.**

Speaker notes:

- The input / output is already in / will be a symbolic form, e.g. mathematical formal language, causal relation graph, etc.
- Goal of reasoning is already given and assumed to be understood.

TODO: draw diagram on perception -> cognition

- **Decision making and ethics.**

Speaker notes:

- As an assistant.
- **Consciousness / self-awareness / active learning.**

Speaker notes: The motivation of seeking explanatory power, creativity, and the desire to reason are hard-coded in the system. In other word, a zombie AI that has no consciousness.

2 Why important at this moment

Outline recap:

1. What's AI2Reason
2. **Why important at this moment**
 - A. Necessity
 - B. Readiness
 - C. Mutual benefit
 - D. Social impact
3. Why is it hard but promising now
4. My Next step




A. Necessity

- AGI should be able to **develop creative** yet still **persuasive** ideas by providing strong reasons to support them.
- Necessity of AI2Reason = necessity of AGI
- Why AGI? Self-improve intelligence ➡ superintelligence !!
- ~~Who knows how far we are from meeting with aliens? As long as we don't destroy ourselves before that ...~~

Speaker notes: I'd like to claim the necessity of AI2Reason by reducing it as the question of necessity of AGI.

- Current LLM **dreams/hallucinates/bullshits**. Because they don't particularly care about the truth. They just care what word will high likely follow and entertain human. We ❤️ LLMs because of its creativity, not because it is a good intelligent system.
- E.g. Ask "show me why $\gcd(n, n-1) = 1$ ":



B. Readiness

- Thanks to the development of ML, particularly LLM, it is now more feasible than ever. We can
 -  **neuralize** many modules via auto-differentials
 -  make use of the infinite expressive power of **natural language**
 -  take LLMs as working (not satisfying) **creative engine**
- Take GPT-4 system as an working example of such system, which isn't doing that bad.

B. Readiness cont'




- Psychologists and philosophers have been studying reasoning for a while.
- Programming logic community have been studying logic for a while.
- Recent progress: TODO

C. Mutual benefit

- Mutual benefit between areas
-  Taking inspiration from theories on reasoning to AI facilitates the development of AGI.
-  At the same time, building computational model is a good way to **complement/connect** current normative/philosophical/explanatory theory and descriptive/psychological understandings. Thus this is a way to characterize what is plausible for such kind a system.

Speaker notes: Connection: by providing implementation of descriptive theories, we can fill in practical gaps. By providing implementation of normative theories, we suggest feasible instantiation or alternatives.

D. Social impact

-  Educational **diagram** of reasoning for future generations.
-  Promote interdisciplinary collaboration. By promoting AI2Reason, we help foster an **environment** where researchers collaborate to advance AI technology more holistically.
-  Positive future for humanity: **advance boundary** of intelligence, shape the future of humanity positively

Speaker notes: - AI2Reason can be an educational diagram for future generations to practice their reasoning skills, as an act to improve humanity. - After all, researchers in academia are motivated by the desire to contribute to advancements that could shape the future of humanity positively.

3 Why is it hard but promising now

Outline recap:

1. What's AI2Reason
2. Why important at this moment
3. **Why is it hard but promising now**

- Human is so smart
- My Point of view
- Under this view, how to frame the problem?
- Mind map
- Which part of it has different situation than it was to be better improved?

4. My Next step

Human is so smart

Human can capture concepts in so little context, mimic rules from so few examples, yet still be able to generalize to genuinely new situations.

Speaker notes: - "The word "five" has the Roman numeral "iv" in it? Show me how you developments your answer."

My Point of view for AI2Reason

- People know LLM sucks in reasoning, and they've tried different heuristic-inspired methods to improve its performance on benchmarks.
- Yet few people sit down and think about what is reasoning. This topic has a long history in philosophy and psychology. Why not learn from them?

My framing of the problem

- Before getting to the next level, the computational model I hope to build right now is an auto-differential neural-symbolic system.

Speaker notes:

- Neural-symbolic system has high inductive bias.
- Although the trend of AI research shows a shift from more symbolic / hand-crafted knowledge to more data-driven / weaker inductive bias, our understanding of reasoning is still in its early stage.
- We need neural-symbolic system as an intermediate solution to have better data efficiency.
- AI4MATH? It is just a play ground. Math is the most abstract and formal yet established language we have. It is the best way to test the reasoning ability of an AI system.
- Automated theorem proving? Again, a play ground that is well-defined and established.
- These leads to my mind map to decompose the problem into different levels of modules.

Mind map

Speaker notes: go the interactive html

Which part of it has a different situation from it was to enable the chancing of being better improved?

- Automated theorem proving is getting more and more attention. Better tools are built.
- Language is powerful. LLMs enable the connection of different modules.

4 My Next step

Outline recap:

1. What's AI2Reason
2. Why important at this moment
3. Why is it hard but promising now

4. My Next step

- An automated theory prover (ATP)

An automated theory prover (ATP):

- that writes proofs in **verifiable** mathematical formal language
- that provides most insightful proofs and the **motivation** of giving these proofs
- **[new]** that can draw inspiration but make analogy between subgoals in hand and **proof flows** of known lemmas

5 Community

Outline recap:

1. What's AI2Reason
2. Why important at this moment
3. Why is it hard but promising now
4. My Next step

5. Community

- People I consider relevant to this direction
- AI2Reason community@Slack

People I consider relevant to this direction

- Yuhuai Tony Wu @xAI: Minerva and autoformalization
- Brenden Lake @NYU: systematicity
- Denny Zhou @Google: CoT stuff
- Noah Goodman @Stanford
- Josh Tenenbaum @MIT
- Jeremy Avigad @CMU
- Kaiyu Yang @Caltech
- Kenneth D. Forbus @Northwestern

- Tom Griffiths @Princeton
- ...

Thank You!

- <https://zoryzhang.notion.site>

- zoryz2@illinois.edu;
- zory_zhang@X;
- zoryzhang@wechat

Join Slack right now to see what exciting things are happening! We welcome everyone who is interested in this direction.

Q&A time!

AI2Reason Community@Slack



