AI2Reason

Build and characterize artificial reasoning system that is truth-seeking, persuasive, and creative.

By Zory Zhang

Outline

Goal: introduce and ask for opinion on my long term vision of AI2Reason.

- ❶ What's AI2Reason
- ❷ Why important at this moment
- ❸ Why is it hard but promising now
- ❹ My Next step

❶ 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

- ❌ just *build* stronger computational model
- ✓ but **characterize** how to let AI reason in a **truth-seeking, persuasive, and creative** manner.

B. Some key features

- Human reasoning: from everyday problem-solving to scientific innovation.
- Let me examplify.

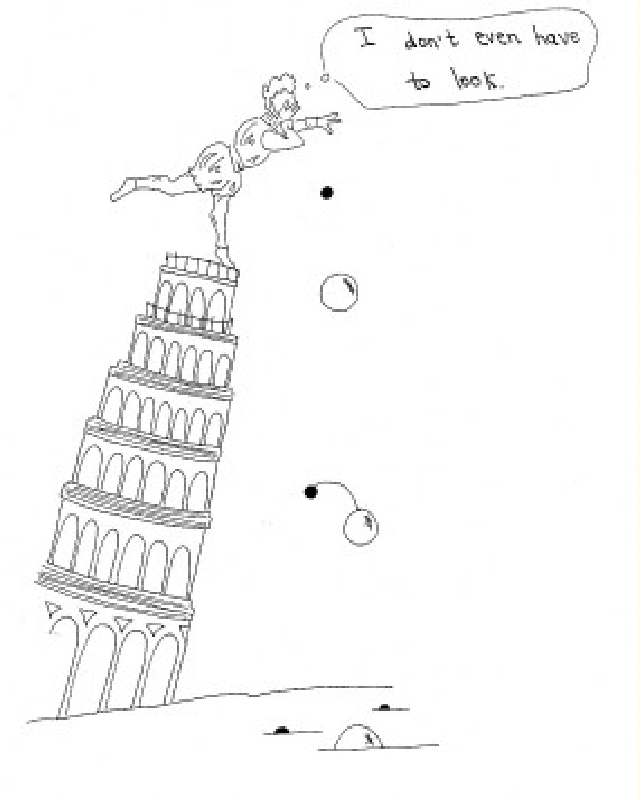
From **solving math word problems**

- Representative problem-solving skill
  - Math: just play ground to study reasoning
  - Formal math language: established play ground

1. ➡️ Formulation: translate into formal language
2. ➡️ Planning: goal decomposition
3. ➡️ Automated reasoning: recursively solve subgoals

To serve as a scientific enquiry assistant

- observer
- hypothesis generation
- reasoning on hypothesis as explanation
- thought experiment / real world experiment

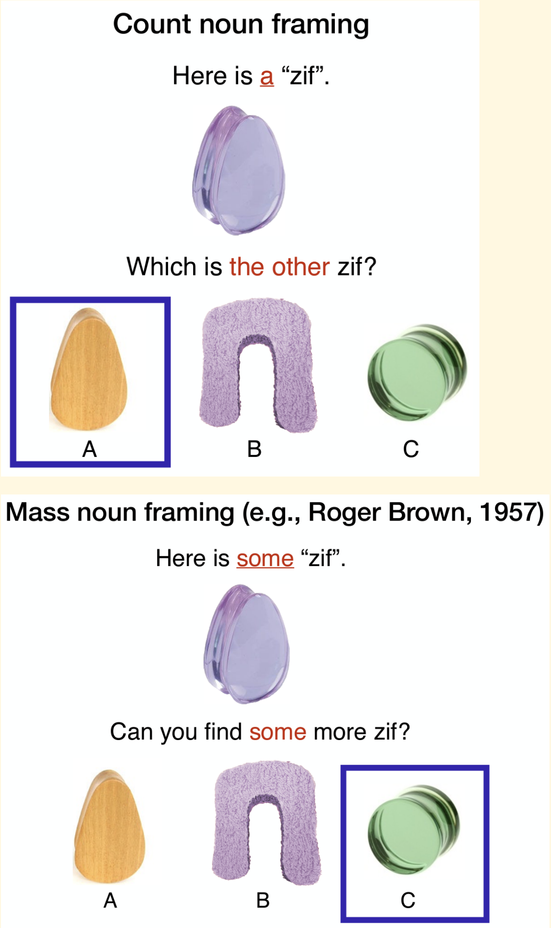


To **develop new theory**

- Scientific concept / diagram innovation
- E.g. weight of object ➡️ universal gravity

C. What aspects of intelligent system are covered?

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



Count noun framing

Here is a “zif”.

Which is the other zif?

Mass noun framing (e.g., Roger Brown, 1957)

Here is some “zif”.

Can you find some more zif?

D. What aspects of intelligent system are not covered?

- **Perception / visual reasoning / embodied reasoning**.
- **Decision making and ethics**.
- **Consciousness / self-awareness / active learning**.

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

- LLMs **dream/hullucinate/bullshit**. They care about
  - ✔ what word will high likely follow
  - ✔ entertain human
  - ❌ truth
- We ❤ LLMs because
  - ✔ creativity
  - ❌ intelligent system with strong generalization

B. Readiness

- 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**
- GPT-4 system:
  - working example
  - 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.

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

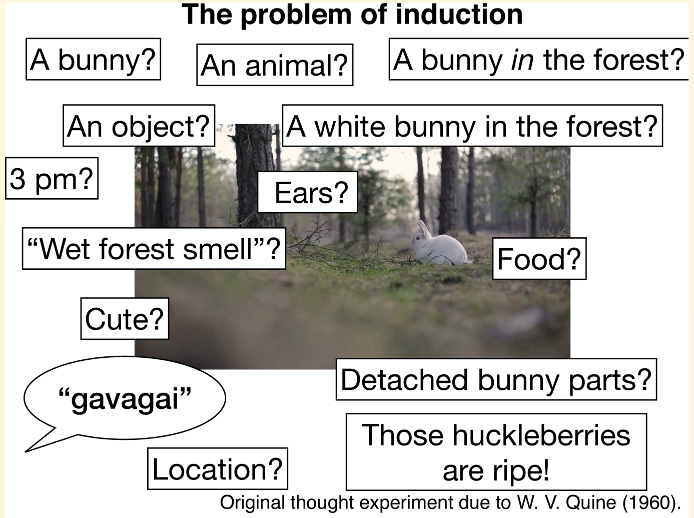
3 Why is it hard but promising now

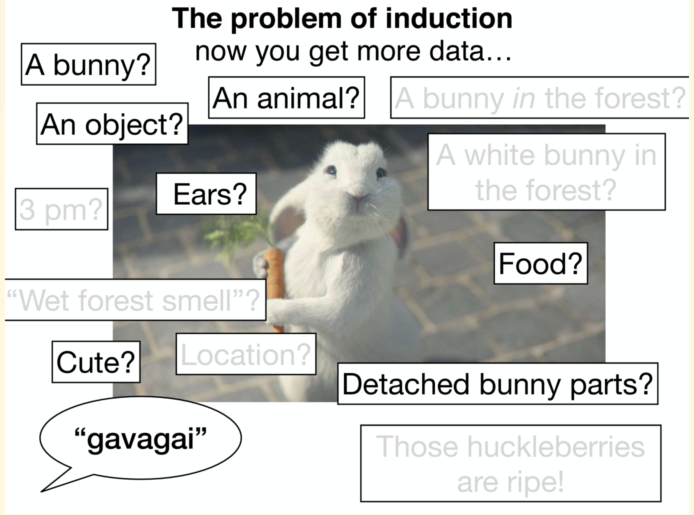
Outline recap:

1. What's AI2Reason
2. Why important at this moment
3. **Why is it hard but promising now**
    - A. Human is so smart
    - B. My Point of view
    - C. Under this view, how to frame the problem?
    - D. Mind map
    - E. Which part of it has different situation than it was to be better improved?
4. My Next step

A. 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.





B. 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?

C. 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.
- 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.

D. My mind map

E. 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. The stronger LLMs become, the better future these is.

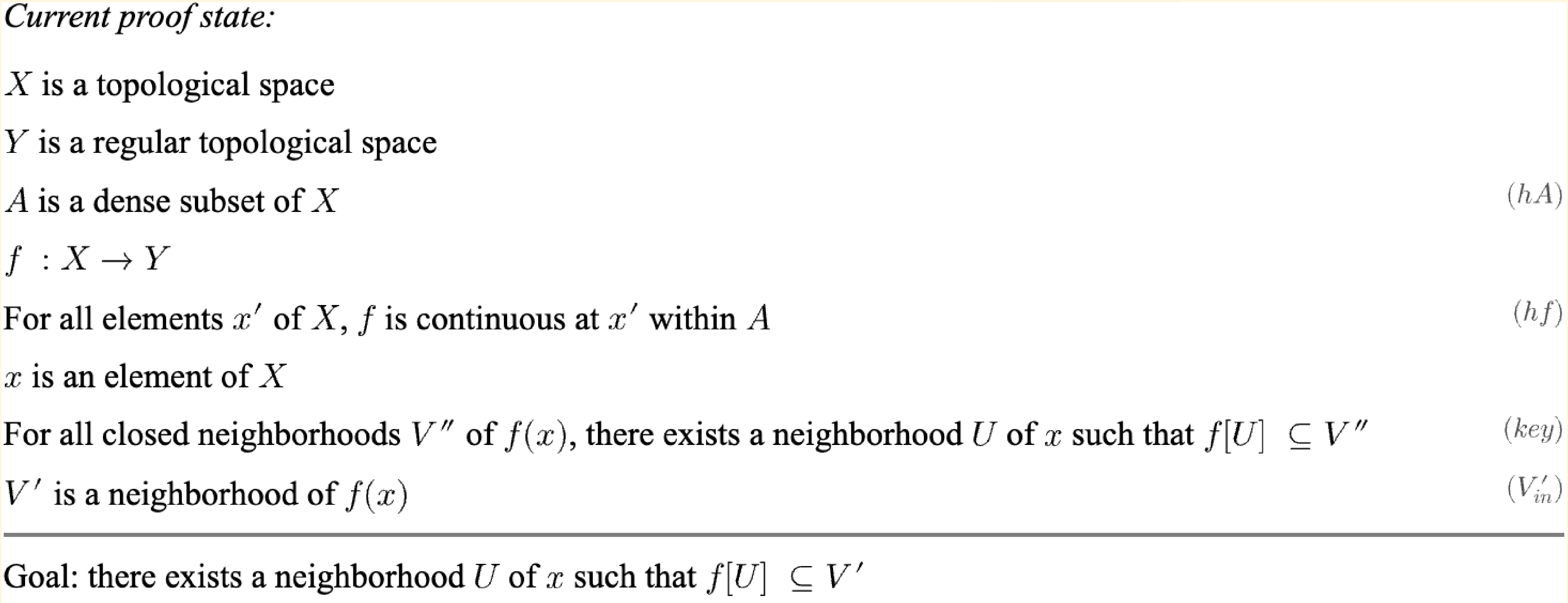
4 My Next step

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) with analogy

An automated theory prover (ATP) with analogy

- 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 bu make analogy between subgoals in hand and **proof flows** of known lemmas



Current proof state:

X is a topological space

Y is a regular topological space

A is a dense subset of X (hA)

$f \ : X \to Y$

For all elements x' of X, f is continuous at x' within A (hf)

x is an element of X

For all closed neighborhoods V'' of $f(x)$, there exists a neighborhood U of x such that $f[U] \ \subseteq V''$ (key)

V' is a neighborhood of $f(x)$ (V'_{in})

Goal: there exists a neighborhood U of x such that $f[U] \ \subseteq V'$

A demo by Patrick Massot

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 highly relevant to this direction
   - Me >_<

People I consider highly 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: LeanDojo
- Kenneth D. Forbus @Northwestern
- Tom Griffiths @Princeton
- ...

Me >_<

- 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.

Thank You! Q&A time!

AI2Reason Community@Slack