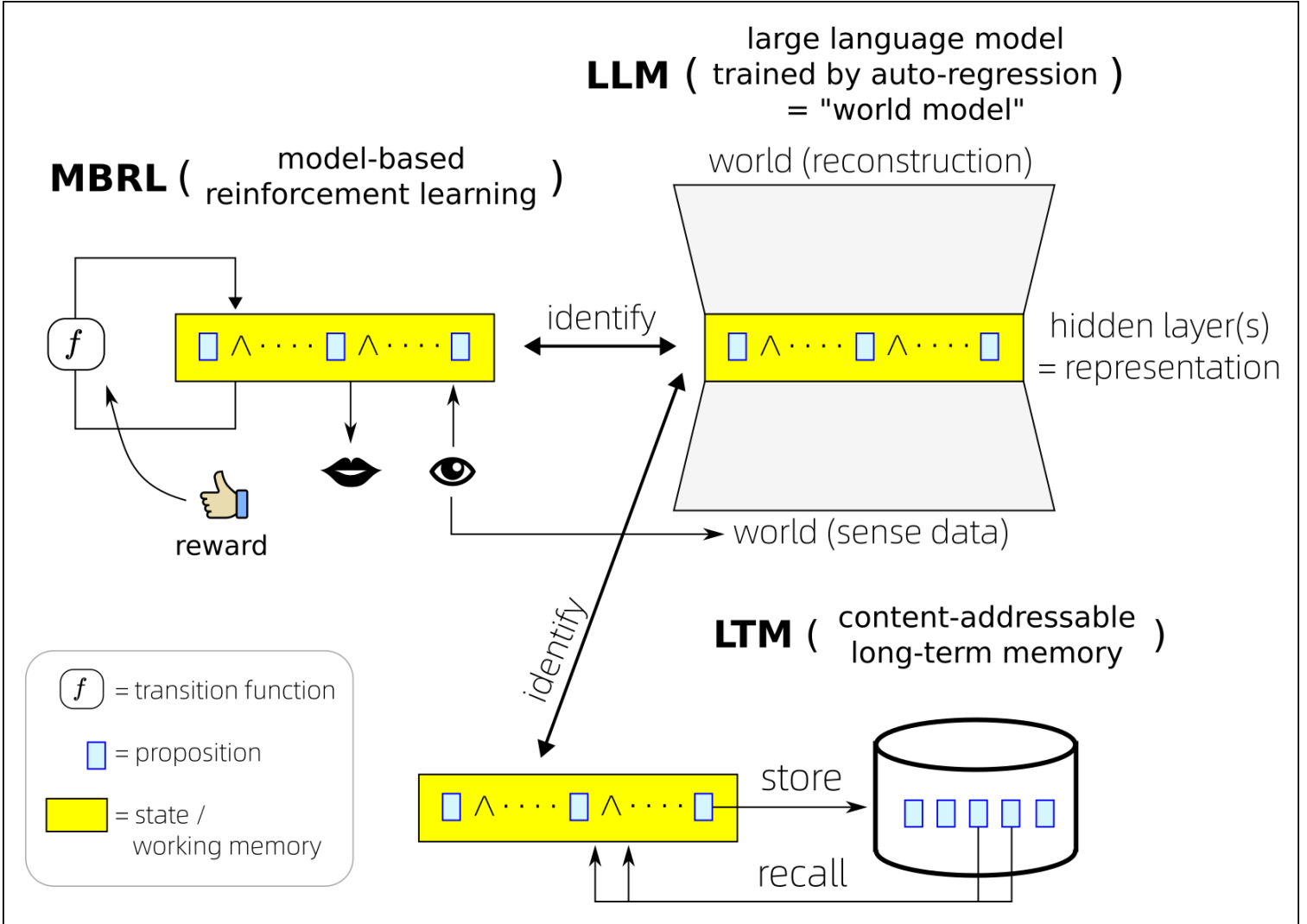# Global AGI Project

About Myself

- Hello. This is me, some years ago, in HKU library:

- You might have noticed I have been applying to every cohort of this program.

- But this is the first time I pitch on the business idea that is my true passion, which is AGI.

- In the past attempts, I have tried to disguise AGI under other business ideas because somehow I wanted to approach AGI from an oblique angle, to perhaps squeeze into your incubation program. And also because I don't want to pitch an idea that sounded like blue-sky research or that I'm out of touch with reality. But this is no longer necessary. The time is now ripe for AGI.

- I have started independent research on AGI in earnest, since 2004. During the past 20 years there was never a day I did not work on AGI.

- Hong Kong is a technologically very backwards place (not in the sense that people do not own slick cell phones, but we are not part of the developers of these technologies). Perhaps this is a too-harsh criticism, since basically the rest of the world can be said the same, except for USA, which dominates most high tech industries. For a long time I could not find any partners or sponsors to work on AGI. Not only that I received no help, but often sarcasm and bitterness from my fellow citizens as well.

- Maybe the number of people interested in AGI in HK is not exactly zero, but if this number is too small, then practically I really cannot find anyone with a similar interest. I have cried over this every day for many years. Perhaps, HKAI Lab has a bit of a responsibility to provide an opportunity to draw such talents together?
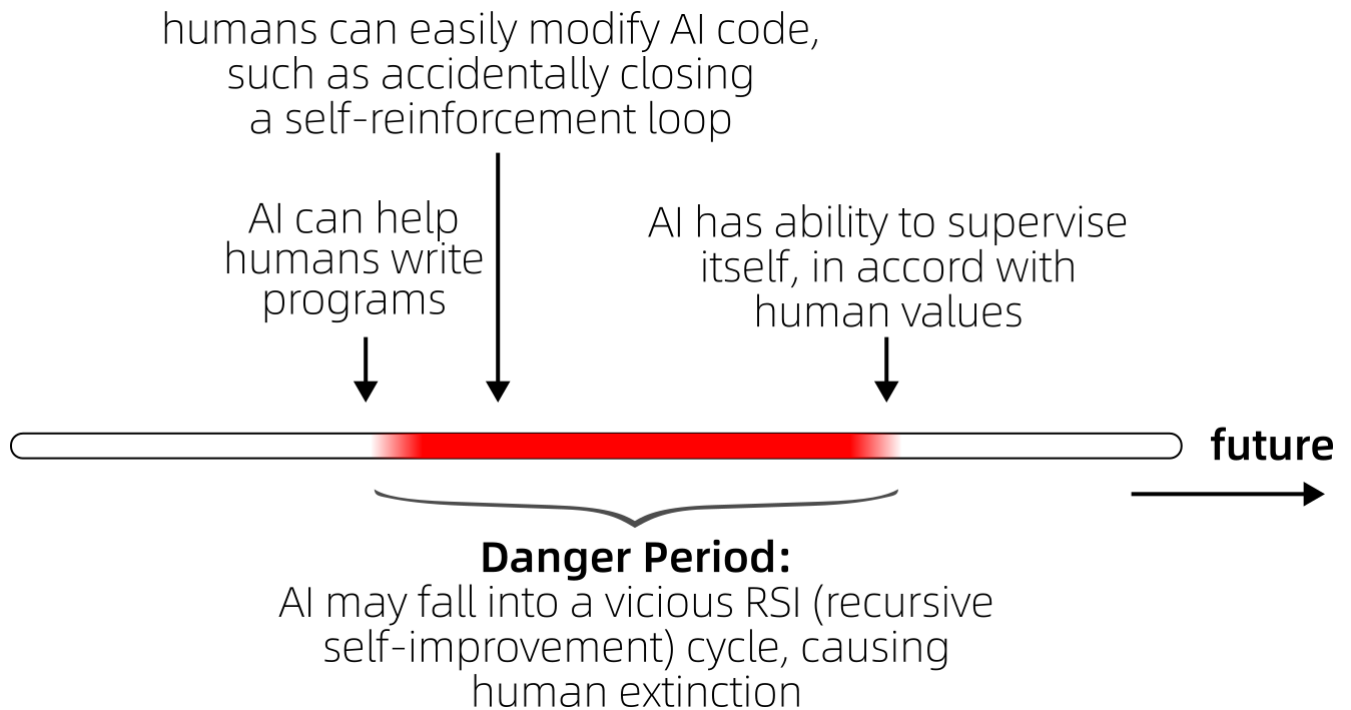
AGI

- Assuming the reader is not an expert in AGI, I might explain it this way: I think the development of AGI has reached a stage now where "bottlenecks" no longer exist. In the past there were so-called "bottlenecks" in the sense that theoretical obstacles existed for which we cannot foretell their solutions. But such obstacles no longer exist and we're at a stage of just engineering AGI systems, using tools and techniques that has been demonstrated to work and are reasonably well-understood.

- At this point it is important to design a cognitive architecture where researchers can clearly understand the internal workings of an AGI, for example, what is "working memory" and where is it located in the architecture, etc.

- I have independently proposed the ideas that LLMs (large language models) are "world models" in model-based reinforcement learning, and that Transformers can be seen as performing the function of logic rule-bases. These are the core ingredients of an AGI. Moreover, I proposed a way to treat "thinking" as "actions" in mental space, unifying thinking and acting under reinforcement learning. My theories have been extensively published on 知乎.com (in addition to the AGI International Conference) which attracted a moderate following. For example: AGI 架构综述

- Our basic architecture is as follows: (the long-term memory part has been demonstrated to work in Google's recent paper "*Memorizing Transformers*")

Why Global?

- There is a widely recognized danger of AGI getting too strong and causing human catastrophes:

humans can easily modify AI code,
such as accidentally closing
a self-reinforcement loop

AI can help
humans write
programs

AI has ability to supervise
itself, in accord with
human values

future

Danger Period:
AI may fall into a vicious RSI (recursive
self-improvement) cycle, causing
human extinction

- The only sensible solution for this problem seems to be to allow a diversity of AGIs trained with different value systems by humans.

- There is no operable definition that can distinguish between human emotions and machine emotions – they are just reward functions in reinforcement learning – implying that as AGI attains higher cognitive powers (including self-consciousness) they could become indistinguishable from human beings. It is important to allow humans to represent their culture and values globally and adequately.

- Currently the USA (or Anglo-American companies) dominates the global technological landscape, especially in the field of AI. This may pose a threat to diversity if there is a systematic bias towards Anglo-American values. Another weakness may come from the homogeneity of AGI architectures designed by similar companies. It would be beneficial for humanity as a whole if there is a globally collaborative AGI project.

- Hong Kong's special position at the intersection of East and West makes it an ideal place to launch this global AGI project.

Our Team / Current State of Our Project

- 158 people on WeChat group, mostly students and researchers of AGI, mainly from mainland China

- An international team is also in the process of forming, on Telegram and Discord. (I used to have an international AGI research team circa 2000's)

- 1-2 angel investors (from mainland China)

- AGI prototype planned to complete training in mid-May this year (2023). The main point of the demo is integrating reinforcement learning with LLM according to our AGI theory.

- We are forming partnerships (indeed, an ecosystem) in business applications of AGI, such as simple businesses utilizing GPT-like services, AIGC, and Web3 technologies (our own AGI project is operated as a DAO – distributive autonomous organization).