AGI Architecture

YKY [November 2, 2025]

This is the "standard model" of AGI that I proposed about a decade ago and seems to have coincided with other mainstream research group's thinking (probably independently) such as RLHF and DeepSeek.

It is based on Richard Sutton's framework of reinforcement learning (RL) and later integrated with large language models (LLMs) as the latter becomes hugely successful.

I call it the "standard model" as it has a very firm theoretical foundation and seemed to me at one point to be inevitable for anyone interested in building AGI or strong AI. That is, as far as you're interested in AGI as a tool for humans to solve general problems, and not as a sort of "god" to reign over humans. Even this is sufficient to disrupt our existence so much that it will probably lead to a "post-human" future.

Without further ado.... RL is a

- Here

0   This

- Here

①

1 This

- Here

②

2 This

- Here