

Ethical and Safe Artificial General Intelligence (AGI)

(Including Scenarios with Technology from Meta, Amazon, Google, DeepMind, YouTube, TikTok, Microsoft, OpenAI, Twitter, Tesla, Nvidia, Tencent, Apple, and Anthropic)

Provisional Patent Application – March 17, 2023 -- Inventor: Craig A. Kaplan

ABSTRACT

This invention enables ethical and safe Artificial General Intelligence (AGI) to emerge from a network of human and AI problem solvers. The AI problem solvers – Advanced Autonomous Artificial Intelligences (AAIs) – are customized by individual users to accomplish tasks and/or earn money on behalf of users.

Because human problem solvers contribute expertise that the AAIs lack, enabling the network to perform as an AGI on “Day One”, this approach is the fastest path to AGI. Because humans are in the loop, providing ethical instruction from the beginning, and because the system and methods include scalable ethical checks, this invention is the safest path to AGI.

Over time, humans do less of the intellectual work as the AAIs learn both intellectual skills and values/ethics from the humans. Eventually, the AAIs are doing almost all intellectual activity faster and better than the humans, while the humans are still providing ethical guidance. Different learning/training/tuning methods are explained including approaches beyond the standard Transformers and Deep Learning techniques.

The preferred implementations of the invention leverage the data, products, and platforms of other technology companies – including, without limitation, Meta, Amazon, Google, DeepMind, YouTube, TikTok, Microsoft, OpenAI, Twitter, Tesla, Nvidia, Tencent, Apple, and Anthropic – in order to customize AAIs more quickly and powerfully than would otherwise be possible. A simple implementation that can be realized, with or without participation of potential partner companies, is illustrated and discussed.