World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:18, No:06, 2024

A Collective Intelligence Approach to Safe Artificial General Intelligence

Authors : Craig A. Kaplan

Abstract : If AGI proves to be a “winner-take-all” scenario where the first company or country to develop AGI dominates, then the first AGI must also be the safest. The safest, and fastest, path to Artificial General Intelligence (AGI) may be to harness the collective intelligence of multiple AI and human agents in an AGI network. This approach has roots in seminal ideas from four of the scientists who founded the field of Artificial Intelligence: Allen Newell, Marvin Minsky, Claude Shannon, and Herbert Simon. Extrapolating key insights from these founders of AI, and combining them with the work of modern researchers, results in a fast and safe path to AGI. The seminal ideas discussed are: 1) Society of Mind (Minsky), 2) Information Theory (Shannon), 3) Problem Solving Theory (Newell & Simon), and 4) Bounded Rationality (Simon). Society of Mind describes a collective intelligence approach that can be used with AI and human agents to create an AGI network. Information theory helps address the critical issue of how an AGI system will increase its intelligence over time. Problem Solving Theory provides a universal framework that AI and human agents can use to communicate efficiently, effectively, and safely. Bounded Rationality helps us better understand not only the capabilities of SuperIntelligent AGI but also how humans can remain relevant in a world where the intelligence of AGI vastly exceeds that of its human creators. Each key idea can be combined with recent work in the fields of Artificial Intelligence, Machine Learning, and Large Language Models to accelerate the development of a working, safe, AGI system.

Keywords : AI Agents, Collective Intelligence, Minsky, Newell, Shannon, Simon, AGI, AGI Safety
Conference Title : ICAGIC 2024 : International Conference on Artificial General Intelligence and Consciousness
Conference Location : San Francisco, United States
Conference Dates : June 03-04, 2024