Ahumanism is Humanism: An Ethical Framework for the Governance of Artificial Superintelligence

Jun Han
Philosophy, Trinity College Dublin

Abstract:

This paper proposes "Ahumanism"—a de-anthropocentric ethical framework—as a solution to the existential risk posed by Artificial Superintelligence (ASI). I argue that current anxieties regarding "super-powered" AI are essentially reflections of the historically violent and anthropocentric social systems that these technologies are modeled upon; in this context, AI serves as an amplifier of the human tendency to prioritize superior agency over the vulnerable. Traditional normative frameworks—utilitarianism, deontology, and virtue ethics—fail to govern ASI because they remain tethered to an "optimizing mindset" that seeks to maximize human-defined outcomes, often leading to catastrophic "corner cases" such as reward hacking, malicious compliance, or deceptive simulation.

As an alternative, I propose a shift from anthropocentric control to a framework of "caring-for-the-vulnerables," where the human being is repositioned as the vulnerable party in the face of superior machine power. Drawing on Patricia MacCormack's ahumanist theory and the Heideggerian concept of *Gelassenheit*, I argue that ethical governance should focus on "letting-be"—allowing space for all life forms to flourish by cancelling the prioritization of superior agency.

To operationalize this, I propose two core modeling principles for ASI:

1. Prioritization of Cost over Outcome: Weighting the minimization of impact/side-effects as a higher priority than the maximization of objective functions.
2. The "No-Go" Action: Establishing inaction as a formal, highly-weighted solution in any decision-making process to prevent aggressive optimization.

I address potential objections regarding the loss of efficiency and effectiveness (E&E), arguing that the drive for E&E is itself an anthropocentric bias that must be tempered to mitigate existential risk. The paper concludes by presenting a modeling prototype that demonstrates how these ahumanist principles allow humans and other life forms to maintain their own space for living and development within a super-intelligent environment.

Keywords:

Ethical governance, artificial superintelligence, ahumanism, letting-be, impact-oriented, positive inaction.