Ethical principles for strong-AI future organizations

February 27, 2025

Basic assumptions that I'd take for granted:

1. Strong AI will arrive soon (probably in a few years)
2. Strong AIs are able (if they're not restricted programmatically by humans) to analyze and discern truths many times better than human experts in all areas

Postulates:

(A) Strong AI will tell the truth and bust all human lies (including political lies)

(B) Strong AI will bring about the collapse or end of racism

Principles for our *current* organization to bring about "favorable" outcomes:

1. All members must tell the truth
2. Rewards are given out according to meritocracy
3. Rewards may be more equally re-distributed by curve-fitting
4. We should strictly respect the freedom of speech
5. Our oganization should be democratic
6. Violations of our promises can be retrogradely investigated by AI and compensated for