Related

Scratchpad

Outline

This depends on how powerful it is, and the level of our alignment ability. Different people have suggested different goals depending on what they think is the most likely kind of first AGI system.

1. If there are many weak AGI systems, then there could be a variety of different specific goals. Such as in Drexler's [Comprehensive AI Services](#).
2. If we only know how to align a relatively weak goal safely, we might instruct it to take a pivotal act in order to buy time to learn how to align more powerful systems
3. Try to align it towards a human category like [human flourishing](#)
4. If we are able to completely align an arbitrarilay powerful system we might want to give it the objective of implementing [CEV](#)
5. [Helping with alignment](#) .
    {Each of these need expansion and maybe more links}