AGI: Racing and cooperating

Seán Ó hÉigeartaigh

Centre For the Study of Existential Risk &

Leverhulme Centre for the Future of Intelligence

Centre for the Study of Existential Risk
“Whoever becomes the leader in this sphere will become the ruler of the world “

“To seize the major strategic opportunity for the development of AI, to build China’s first-mover advantage in the development of AI “

Racing and cooperating
’Races’ in contemporary AI

- National strategies
- Competition for research talent
- Industry and business application
- Emerging markets
- Competition can drive innovation
'Races' in contemporary AI

- Concentration of power
- Military applications
- Many different competitions with different (changing) leaders
Emergence of an AGI race

- AGI currently speculative future prospect
- At some point may become imminent-seeming
- A different race dynamic?
  - No room for many winners?
  - Unassailable leadership achievable?
- Beliefs matter
Concerns around a race to AGI

- Safety
- Fair and beneficial deployment
- Expertise-sharing and cooperation
- Tensions and conflict
- What if race is won?
In an ideal world?

- Globally cooperative venture
- Safe and controlled
- Global stakeholders convinced that benefits to be distributed worldwide
What can we do now?

- Rhetoric
- Cooperation
  - Sharing and open-ness models
- Global engagement
- Precedents

Long-Term Safety

We are concerned about late-stage AGI development becoming a competitive race without time for adequate safety precautions. Therefore, if a value-aligned, safety-conscious project comes close to building AGI before we do, we commit to stop competing with and start assisting this project. We will work out specifics in case-by-case agreements, but a typical triggering condition might be “a better-than-even chance of success in the next two years.”
Topics and questions

- Models for a joint AGI development project
- Models for an global AGI benefits agreement
- What would be ‘good enough’?
- Alternatives to an ‘arms race’ framing?
- How do we talk about an AGI race?