The Comprehensive AI Services framework

Eric Drexler

Future of Humanity Institute
University of Oxford

Beneficial AGI Workshop
3 January 2019
Puerto Rico

> FHI TECHNICAL REPORT <

Reframing Superintelligence

Comprehensive AI Services as General Intelligence

K. Eric Drexler

Technical Report #2019-1

Some basic questions

Where do AI systems come from? — *Research and development*

What do AI systems do today? — *Perform tasks (i.e., provide services)*

What will AI systems be able to do? — *Automate human tasks (and more)*

What is research and development? — *A bunch of tasks to automate*

Where does AI R&D automation lead? — *Recursive <u>technology</u> improvement*

Where does *that* lead? — ***<u>Comprehensive AI services (CAIS)</u>***

Automated service development *is* general intelligence

The "G" in CAIS does the work of "G" in AGI —

Mostly *descriptive*, partly *prescriptive*

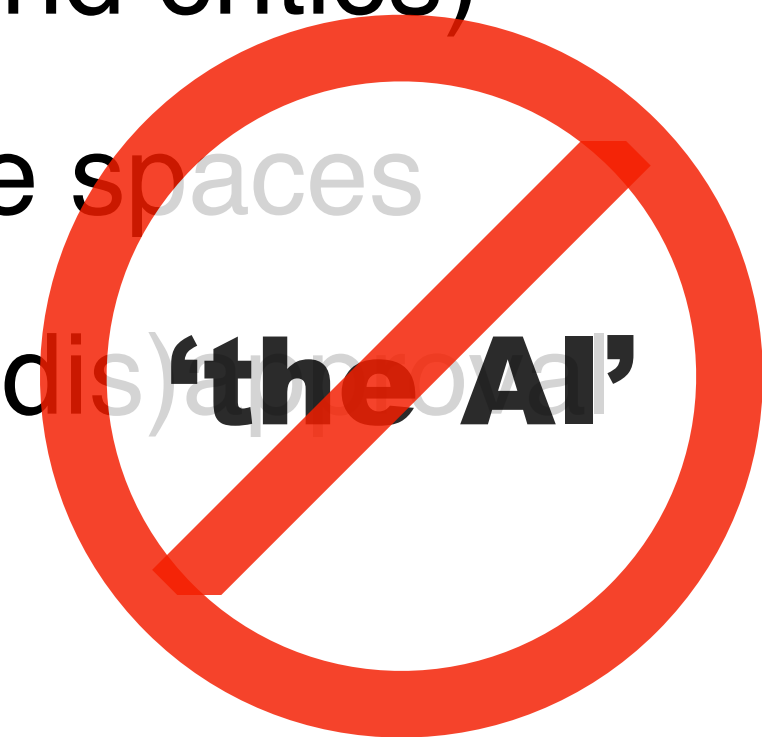| Characteristics	…in AGI-agent models	…in AI-service models
Typical embodiments:	Distinct AI agents	Networked AI services
Action evaluation:	Global utility function	Task objective function
Default goals:	Resources, autonomy	Task completion
Stronger optimization:	More powerful, general	More effective, focused
Recursive improvement:	Self-improving agent	Automated AI R&D
Takeoff threshold:	Agent capabilities	Technological capabilities
Takeoff dynamics:	Autonomous, abrupt	Managed, potentially swift
Developer experience:	Unprecedented, novel	Extensive, incremental
Autonomous AGI:	Uniquely valuable	Needlessly risky

Safety-relevant affordances:

- Implementation relationships (A optimizes B to provide service C)

- Optimization as a constraint (components perform known tasks)

- Separation of learning and performance (stable competencies)

- Competition among systems (enables comparision, choice)

- Adversarial systems (red-team/blue-team, proposers and critics)

- Matching tasks to services in semantically interpretable spaces

- Employing predictive models of human concerns and (dis)approval 'the AI'

Descriptive and prescriptive…

Description:

Consider <u>patterns</u> of system development and structure

Prescription:

Exploit <u>affordances</u> of system development and structure