Define Artificial General Intelligence!

Levels of AGI: Operationalizing Progress on the Path to AGI

John I Davies

It is widely anticipated that the first steps our species takes into interstellar space will use some form of what is loosely called artificial intelligence. This begins at least as long ago as Arthur C Clarke's *Profiles of the Future* - where he suggests that, for interstellar space "...it may be that only creatures of metal and plastic can ever really conquer it, as they have already started doing." But can these "creatures" ever exhibit artificial general intelligence (AGI) rather than the limited AI we have so far?

Google DeepMind [1] wants to define what counts as artificial general intelligence, as reported in MIT Technology Review - *Google DeepMind wants to define what counts as artificial general intelligence*. AGI is one of the most disputed concepts in technology. These researchers want to fix that.
The paper is *Levels of AGI: Operationalizing Progress on the Path to AGI* (arxiv.org/abs/2311.02462), Meredith Ringel Morris et al (Google DeepMind).
They propose a framework for classifying the capabilities and behaviour of Artificial General Intelligence (AGI) models and their precursors by defining levels of AGI performance, generality, and autonomy.
Starting by analysing existing definitions of AGI they distil six principles for a useful nomenclature [2] for AGI to satisfy -

1. Focus on capabilities, not processes. This allows the researchers to ignore issues they regard as processes including - systems think or understand in a human-like way, or systems possessing qualities such as consciousness or sentience. Thus they put aside some of the tougher philosophical questions about AGI.
2. Focus on Generality and Performance, arguing that both generality and performance are key components of AGI.
3. Focus on Cognitive and Metacognitive Tasks. But they suggest that the ability to perform physical tasks increases a system's generality, but should not be considered a necessary prerequisite to achieving AGI - ie no robots required!
4. Focus on Potential, not Deployment since requiring deployment as a condition of measuring AGI introduces non-technical hurdles such as legal and social considerations, as well as potential ethical and safety concerns.
5. Focus on Ecological Validity by choosing tasks that align with real-world (ie ecologically valid) tasks that people value (construing "value" broadly, not only as economic value but also social value, artistic value, etc). In other words AGI must meet everyday human standards of "General Intelligence".
6. Focus on the Path to AGI, not a Single Endpoint. They cite the adoption of a standard set of Levels of Driving Automation [3] allowed for clear discussions of policy and progress relating to autonomous vehicles and suggest there is value in defining "Levels of AGI."

[1] Demis Hassibis, a founder of DeepMind, famously said that his mission was to "solve intelligence, and then use that to solve everything else" www.theguardian.com/technology/2016/feb/16/demis-hassabis-artificial-intelligence-deepmind-alphago
[2] They use the word "ontology" whose meaning is not clear in this context. Its more usual meaning is the meaning of terms in context.
[3] *SAE International. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, April 2021 www.sae.org/standards/content/

They believe this delimited process nevertheless maps onto goals for, predictions about, and risks of AI. The paper includes nine case studies, proposed definitions of AGI, which they review critically as a foundation for their proposed definition. They begin with the Turing Test [1] - agreeing with Turing by saying that whether a machine can "think," while an interesting philosophical and scientific question, seems orthogonal to the question of what the machine can do; the latter is much more straightforward to measure and more important for evaluating impacts. Hence their proposal that AGI should be defined in terms of capabilities rather than processes. They go through a number of distinguished thinkers ending with the recent development of large language models (LLMs) which have been claimed to fit some definitions of AGI.

They present a table displaying a levelled, matrixed approach toward classifying systems on the path to AGI based on depth (performance) and breadth (generality) of capabilities. They note that general systems that broadly perform at a level N may be able to perform a narrow subset of tasks at higher levels.

They remark that -

> "Competent AGI" level, which has not been achieved by any public systems at the time of writing, best corresponds to many prior conceptions of AGI, and may precipitate rapid social change once achieved."

This is a simplified version of their table -

Performance (rows) x Generality (columns)	**Narrow: clearly scoped task or set of tasks**	**General: wide range of non-physical tasks, including metacognitive abilities like learning new skills**
Level 0: No AI	Narrow Non-AI calculator software; compiler	General Non-AI human-in-the-loop computing, eg Amazon Mechanical Turk
Level 1: Emerging equal to or somewhat better than an unskilled human	Emerging Narrow AI GOFAI (good old-fashioned AI) ; simple rule-based systems	Emerging AGI ChatGPT, Bard, Llama 2 , Gemini
Level 2: Competent at least 50th percentile of skilled adults	Competent Narrow AI Toxicity detectors such as Jig-saw ; Smart Speakers such as Siri (Apple), Alexa, or Google Assistant.	Competent AGI not yet achieved
Level 3: Expert at least 90th percentile of skilled adults	Expert Narrow AI spelling & grammar checkers such as Grammarly and generative image models such as Imagen	Expert AGI not yet achieved
Level 4: Virtuoso at least 99th percentile of skilled adults	Virtuoso Narrow AI Deep Blue, AlphaGo.	Virtuoso AGI not yet achieved
Level 5: Superhuman outperforms 100% of humans	Superhuman Narrow AI AlphaFold, AlphaZero , StockFish	Artificial Superintelligence not yet achieved

They suggest challenging requirements for future benchmarks to quantify the behaviour and capabilities of AGI models against these levels and discuss how these levels of AGI interact with deployment considerations such as autonomy and risk.

This is just a brief introduction to a paper which may be a useful addition to the debates about AGI, and its narrowing of the definition is a practical step to assist wider judgements of the immediate implications for our species. But the wider scientific and philosophical problems remain and Principium will pay attention to both the narrow and the wide view of AGI with, of course, a specific focus on the implications for our potential interstellar future.

[1] Computing Machinery and Intelligence, A M Turing, Mind (1950) www.cs.colostate.edu/~howe/cs440/csroo/yr2015fa/more_assignments/turing.pdf