Evolving definitions of powerful AI

Please comment or email (jaswsunny at gmail dot com) if something is wrong/missing!

Subtypes of definitions:

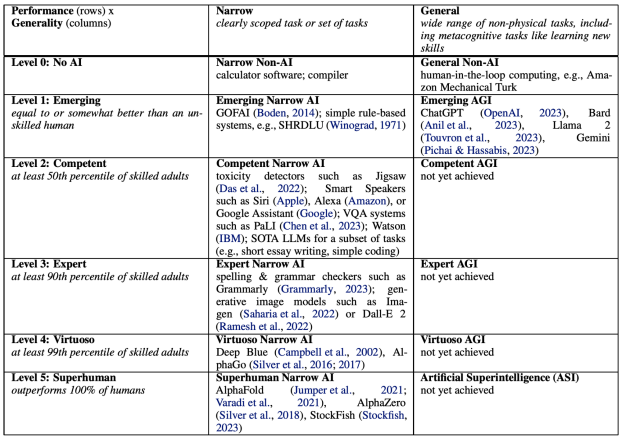
- Processes: Does AI think in "human-like" ways?
- Generality: Can AI learn new tasks it hasn't seen?
- Performance: Can AI complete various tasks?
- Impacts: Will the AI have a big impact on society?

Year	Person	Definition and description	Link and context	Subtype
1947	Alan Turing	"A good working rule is that the ACE can be made to do **any job that could be done by a human** computer, and will do it in one ten-thousandth of the time." "I was researching on what might now be described as an [3] investigation of the theoretical possibilities and limitations of digital computing machines. I considered a type of machine which had a central mechanism, and an infinite memory which was contained on an infinite tape. This type of machine appeared to be sufficiently general."	"Lecture to the London Mathematical Society" from the *Collected Works of Alan M. Turing, Volume 1: Mechanical Intelligence* First known concept of a machine intelligent enough to complete human-like tasks, learn from experience, etc. Also discusses automation, training, alignment, hallucination.	Processes, Performance
1955	John McCarthy, Marvin Minsky, Nathaniel Rochester, Claude Shannon	"We propose that a 2 month, 10 man study of **artificial intelligence** be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves."	Proposal for the Dartmouth Summer Research Project on AI First published usage of "artificial intelligence"	
1980	John Searle (philosopher)	"According to **strong AI,** the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states."	Minds, brains, and programs Searle's "Strong AI" hypothesis, which he refutes, suggests that a computer program simulating a mind is also conscious and has its own mental states. (This is not how most AI researchers or theorists would define AI.)	Processes
1997	Mark Avrum Gubrud (physicist)	"By **advanced artificial general intelligence**, I mean AI systems that rival or surpass the human brain in complexity and speed, that can acquire, manipulate and reason with general knowledge, and that are usable in essentially any phase of ... operations where a human intelligence would	Used in context of advanced military technology in article "Nanotechnology and National Security"	Generality, Performance

		otherwise be needed."	First recorded usage of AGI but went unnoticed by AI community	
2002	Ben Goertzel & Shane Legg	"What is meant by **AGI** is, loosely speaking, AI systems that possess a reasonable degree of self-understanding and autonomous self-control, and have the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems that they didn't know about at the time of their creation. A marked distinction exists between practical AGI work and, on the other hand: • Pragmatic but specialized "**narrow AI**" research which is aimed at creating programs carrying out specific tasks like playing chess, diagnosing diseases, driving cars and so forth (most contemporary AI work falls into this category.) • Purely theoretical AI research, which is aimed at clarifying issues regarding the nature of intelligence and cognition, but doesn't involve technical details regarding actually realizing artificially intelligent software."	Goertzel's book (pub. 2005) needed a title. He didn't like "strong AI," and Shane came up with "AGI." https://goertzel.org/who-coined-the-term-agi Goertzel and Legg are credited with popular diffusion of the term AGI.	Processes, Generality, Performance
2005	Ray Kurzweil	"Artificial intelligence permeates our economy. It's what I define as **"narrow" AI**: machine intelligence that equals or exceeds human intelligence for specific tasks… So what are the prospects for **"strong" AI**, which I describe as machine intelligence with the full range of human intelligence?"	Long Live AI / *The Singularity Is Near*	Generality
2005	Nils Nilsson	"Machines exhibiting true **human-level intelligence** should be able to do many of the things humans are able to do. Among these activities are the tasks or "jobs" at which people are employed. I suggest we replace the Turing test by something I will call the "employment test." To pass the employment test, AI programs must be able to perform the jobs ordinarily performed by humans. Progress toward human-level AI could then be measured by the fraction of these jobs that can be acceptably performed by machines." "Rather than work toward this goal of automation by building special-purpose systems, I argue for the development of general-purpose, educable systems that can learn and be taught to perform any of the thousands of jobs that humans can perform."	Nilsson says "general-purpose, educable" systems are the right approach to achieve HLMI Proposes an "employment test" to measure AI capabilities Human-Level Artificial Intelligence? Be Serious!	Performance, Impacts
2007	Shane Legg & Mark Hutter	"**Intelligence** measures an agent's ability to achieve goals in a wide range of environments." **universal intelligence**	First formal definition of "universal intelligence" https://arxiv.org/pdf/0712.3329	Performance

		Bringing all these pieces together, we can now define our formal measure of intelligence for arbitrary systems. Let E be the space of all computable reward summable environmental measures with respect to the reference machine $\mathcal{U}$, and let K be the Kolmogorov complexity function. The expected performance of agent π with respect to the universal distribution $2^{-K(\mu)}$ over the space of all environments E is given by, $$\Upsilon(\pi) := \sum_{\mu \in E} 2^{-K(\mu)} V_\mu^\pi.$$ We call this the *universal intelligence* of agent π.		
2011	Nick Bostrom & Eliezer Yudkowsky	"**Artificial General Intelligence**" (hereafter, AGI) is the emerging term of art used to denote "real" AI (see, e.g., the edited volume Goertzel and Pennachin 2006). As the name implies, the emerging consensus is that the missing characteristic is generality. Current AI algorithms with human-equivalent or -superior performance are characterized by a deliberately-programmed competence only in a single, restricted domain… It is a qualitatively different class of problem to handle an AGI operating across many novel contexts that cannot be predicted in advance."	The Ethics of Artificial Intelligence	Generality, Performance
2014	Vincent C. Müller & Nick Bostrom	"Define a **'high–level machine intelligence' (HLMI)** as one that can carry out most human professions at least as well as a typical human." "We need to avoid using terms that are already in circulation and would thus associate the questionnaire with certain groups or opinions, like "artificial intelligence", "singularity", "artificial general intelligence" or "cognitive system". For these reasons, we settled for a definition that a) is based on behavioral ability, b) avoids the notion of a general 'human–level' and c) uses a newly coined term."	2014 survey of AI expert opinion — HLMI is used to be debiasing Future Progress in Artificial Intelligence: A Survey of Expert Opinion	Performance
2014	Nick Bostrom	**Superintelligence** is defined as "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest."	From *Superintelligence* book	Performance, Impacts
2016	Holden Karnofsky (OpenPhil)	"Roughly and conceptually, **transformative AI** refers to potential future AI that precipitates a transition comparable to (or more significant than) the agricultural or industrial revolution. I also provide (below) a more detailed definition. The concept of "transformative AI" has some overlap with concepts put forth by others, such as "superintelligence" and "artificial general intelligence." However, "transformative AI" is intended to be a more inclusive term, leaving open the possibility of AI systems that count as "transformative" despite lacking many abilities humans have."	Some Background on Our Views Regarding Advanced Artificial Intelligence	Open Philanthropy	Impacts
2016	Katja Grace et al. (FHI)	""**High-level machine intelligence**" (HLMI) is achieved when unaided machines can accomplish every task better and more cheaply than human workers."	Regular survey of AI experts on views on AGI, first in 2016	Performance, Impacts
2018	OpenAI (Sam Altman, et al?)	"OpenAI's mission is to ensure that artificial general intelligence (**AGI**) — by which we mean highly autonomous systems that outperform humans at most economically valuable work — benefits all of humanity."	Defined in the OpenAI charter: https://web.archive.org/web/20180409161852/https://blog.openai.com/openai-charter/	Performance, Impacts

2019	Francois Chollet	“The **intelligence** of a system is a measure of its skill-acquisition efficiency over a scope of tasks, with respect to priors, experience, and generalization difficulty.” ““General intelligence” is not a binary property which a system either possesses or lacks. It is a spectrum, tied to 1) a scope of application, which may be more or less broad, and 2) the degree of efficiency with which the system translate its priors and experience into new skills over the scope considered, 3) the degree of generalization difficulty represented by different points in the scope considered. It is conceptually unsound to set “artificial general intelligence” in an absolute sense (i.e. “universal intelligence”) as a goal.” “The consensus definition of **AGI**, "a system that can automate the majority of economically valuable work," while a useful goal, is an incorrect measure of intelligence. Measuring task-specific skill is not a good proxy for intelligence.”	On the Measure of Intelligence What is ARC-AGI? Chollet has been a critic of other definitions of AGI, arguing that they measure memorization rather than generality / new-skill acquisition	Generality
2022	Yann LeCun (Meta)	“I think the phrase **AGI** should be retired and replaced by "**human-level AI**". There is no such thing as AGI. Even human intelligence is very specialized. We do not realize that human intelligence is specialized because all the intelligent tasks we can think of are task that we can apprehend. But that is a tiny subset of all tasks. The overwhelming majority of tasks are completely out of reach of un-augmented human intelligence.”	LinkedIn post	Performance
2023	Microsoft Research (Sebastien Bubeck et al)	“We use **AGI** to refer to systems that demonstrate broad capabilities of intelligence, including reasoning, planning, and the ability to learn from experience, and with these capabilities at or above human-level.”	Sparks of AGI paper https://arxiv.org/abs/2303.12712	Processes, Performance
2024	Meredith Morris et al (DeepMind)	“Artificial General Intelligence (**AGI**) is an important and sometimes controversial concept in computing research, used to describe an AI system that is at least as capable as a human at most tasks.”	Levels of AGI for Operationalizing Progress on the Path to AGI Emphasizes generality *and* performance—both exceeding human performance on tasks, and adapting to new tasks via learning	Generality, Performance

				
2024	OpenAI	Bloomberg: OpenAI has internally defined five “levels” of AGI **OpenAI Imagines Our AI Future** **Stages of Artificial Intelligence** Level 1 — Chatbots, AI with conversational language Level 2 — Reasoners, human-level problem solving Level 3 — Agents, systems that can take actions Level 4 — Innovators, AI that can aid in invention Level 5 — Organizations, AI that can do the work of an organization Source: Bloomberg reporting	Bloomberg: OpenAI Sets Levels to Track Progress Toward Superintelligent AI (Archive link) (In a 2023 blog post, Sam Altman similarly distinguishes between “initial AGI” and “successor systems”)	Performance, Impact
2024	Dario Amodei (Anthropic)	“I find **AGI** to be an imprecise term that has gathered a lot of sci-fi baggage and hype. I prefer "powerful AI" or "Expert-Level Science and Engineering" which get at what I mean without the hype.” “By **powerful AI**, I have in mind an AI model—likely similar to today’s LLM’s in form, though it might be based on a different architecture, might involve several interacting models, and might be trained differently—with the following properties: ● In terms of pure intelligence, it is smarter than a Nobel Prize winner across most relevant fields – biology, programming, math, engineering, writing, etc. This means it can prove unsolved mathematical theorems, write extremely good novels, write difficult codebases from scratch, etc. ● In addition to just being a “smart thing you talk to”, it has all the	Machines of Loving Grace	Performance

		"interfaces" available to a human working virtually, including text, audio, video, mouse and keyboard control, and internet access. It can engage in any actions, communications, or remote operations enabled by this interface, including taking actions on the internet, taking or giving directions to humans, ordering materials, directing experiments, watching videos, making videos, and so on. It does all of these tasks with, again, a skill exceeding that of the most capable humans in the world. ● It does not just passively answer questions; instead, it can be given tasks that take hours, days, or weeks to complete, and then goes off and does those tasks autonomously, in the way a smart employee would, asking for clarification as necessary. ● It does not have a physical embodiment (other than living on a computer screen), but it can control existing physical tools, robots, or laboratory equipment through a computer; in theory it could even design robots or equipment for itself to use. ● The resources used to train the model can be repurposed to run millions of instances of it (this matches projected cluster sizes by ~2027), and the model can absorb information and generate actions at roughly 10x-100x human speed5. It may however be limited by the response time of the physical world or of software it interacts with. ● Each of these million copies can act independently on unrelated tasks, or if needed can all work together in the same way humans would collaborate, perhaps with different subpopulations fine-tuned to be especially good at particular tasks. We could summarize this as a "country of geniuses in a datacenter"."		
2025	Sam Altman (OpenAI)	"Systems that start to point to AGI* are coming into view, and so we think it's important to understand the moment we are in. **AGI** is a weakly defined term, but generally speaking we mean it to be a system that can tackle increasingly complex problems, at human level, in many fields."	Three Observations Sam blog post is notably different from the OpenAI site definition, "highly autonomous systems that outperform humans at most economically valuable work" "At human level" and "Many fields" also seem quite vague!	Performance