AGI: Definitions and Potential Impacts

Artificial general intelligence (AGI) is commonly defined as a hypothetical AI system that would match or exceed human-level capabilities across the vast majority of cognitive tasks, though definitions vary across sources. Multiple AI companies have stated intentions to build AGI.[1, 2, 3, 4]

An early definition from 1997 characterized AGI as "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 industrial or military operations where a human intelligence would otherwise be needed." This definition emerged in the context of international security.[5]

OpenAI's charter from 2018 defines AGI as "highly autonomous systems that outperform humans at most economically valuable work."[6] A 2023 paper by Google DeepMind outlines five levels of AGI ranging from "Emerging" (matching unskilled humans in non-physical tasks) to "Superhuman" (outperforming 100% of skilled humans). In this framework, current AI systems like ChatGPT and Gemini are classified as "Emerging AGI," having capabilities in a broad range of domains while remaining below median skilled human performance in most domains.[7]

Some definitions focus on economic effects rather than capabilities, such as generating $100 billion in profits[8] achieving 10% world GDP growth rates.[9] A related concept is *transformative AI*, which emphasizes societal change comparable to the industrial revolution.[10] In the remainder of

---

[1] OpenAI, "OpenAI Charter," April 2018, https://openai.com/charter/.
[2] Anca Dragan, Rohin Shah, Four Flynn and Shane Legg, "Taking a responsible path to AGI," https://deepmind.google/discover/blog/taking-a-responsible-path-to-agi/.
[3] Sean Michael Kerner, "Elon Musk reveals xAI efforts, predicts full AGI by 2029," *VentureBeat*, https://venturebeat.com/ai/elon-musk-reveals-xai-efforts-predicts-full-agi-by-2029/.
[4] DeepSeek, "deepseek-ai," *Hugging Face*, https://huggingface.co/deepseek-ai.
[5] Mark Gubrud (November 1997), "Nanotechnology and International Security," in *Fifth Foresight Conference on Molecular Nanotechnology*, November 1997, archived at https://web.archive.org/web/20110529215447/http://www.foresight.org/Conferences/MNT05/Papers/Gubrud/.
[6] OpenAI, "OpenAI Charter," April 2018, https://openai.com/charter/.
[7] Meredith Ringel Morris, Jascha Sohl-Dickstein, Noah Fiedel, Tris Warkentin, Allan Dafoe, Aleksandra Faust, Clément Farabet, and Shane Legg, "Levels of AGI for Operationalizing Progress on the Path to AGI," in Proceedings of the International Conference on Machine Learning (ICML 2024), arXiv preprint arXiv:2311.02462 (2024), https://arxiv.org/abs/2311.02462.
[8] Maxwell Zeff, "Microsoft and OpenAI Have a Financial Definition of AGI: Report," *TechCrunch*, December 26, 2024, https://techcrunch.com/2024/12/26/microsoft-and-openai-have-a-financial-definition-of-agi-report/.
[9] Siddharth Jindal, "True AGI Means 10% Economic Growth, Says Satya Nadella," *Analytics India Magazine*, February 20, 2025, https://analyticsindiamag.com/ai-news-updates/true-agi-means-10-economic-growth-says-satya-nadella/.
[10] Ross Gruetzemacher and Jess Whittlestone, "Defining and Unpacking Transformative AI," arXiv preprint, https://arxiv.org/vc/arxiv/papers/1912/1912.00747v1.pdf.

this document, we will use AGI to refer to AI that automates the vast majority of non-physical work at an expert level, including complex, months-long projects.

Debate remains about how or when AGI will be achieved. In recent years, large language models (LLMs) and multimodal models have substantially improved through *scaling* – training with more compute, data, and parameters[11] – and through algorithmic improvements.[12] Some believe a few more years of scaling will achieve AGI, while others believe that fundamentally new approaches are needed to reach AGI.[13] One possible accelerating factor is the potential for pre-AGI systems to substantially automate AI research and development (AI R&D), or the process of discovering and implementing improvements to AI capabilities.[14]

Potential impacts

The potential impacts of AGI remain uncertain but could involve transformative benefits and risks.

AGI could greatly increase the intellectual labor available to solve challenges in healthcare, medical research, education, energy technology,[15] and scientific innovation. It would be capable of automating cognitive tasks across domains including software development, cybersecurity, legal analysis, finance, consulting, clinical diagnosis and advising, drug discovery, product design, marketing, creative arts, and engineering.

The impact on employment and wages remains contested. AGI may or may not depress wages or cause mass unemployment, depending on adoption rates, the speed of capability advances, and regulatory barriers to automation. A crucial question is whether demand for human labor can be sustained despite an abundant supply of AGI "workers".[16] AGI development could lead to superintelligence more capable of intellectual tasks than 100% of skilled humans,[17] or advanced

---

[11] Veronika Samborsska, "Scaling up: how increasing inputs has made artificial intelligence more capable," *Our World in Data*, January 2025, https://ourworldindata.org/scaling-up-ai.
[12] Anson Ho et al., "Algorithmic Progress in Language Models," *Epoch AI*, https://epoch.ai/blog/algorithmic-progress-in-language-models.
[13] Association for the Advancement of Artificial Intelligence, "AAAI 2025 Presidential Panel on the Future of AI Research," March 2025, https://aaai.org/wp-content/uploads/2025/03/AAAI-2025-PresPanel-Report-Digital-3.7.25.pdf.
[14] Hjalmar Wijk et al., "RE-Bench: Evaluating frontier AI R&D capabilities of language model agents against human experts," in Proceedings of the International Conference on Machine Learning (ICML 2025), arXiv preprint, https://arxiv.org/abs/2411.15114.
[15] Ricardo Vinuesa et al., "The role of artificial intelligence in achieving the Sustainable Development Goals," *Nature Communications*, January 13, 2020, https://www.nature.com/articles/s41467-019-14108-y. (Note: Sources that refer to AI but not AGI can also provide some insight into potential impacts of AGI.)
[16] Anton Korinek and Donghyun Suh, "Scenarios for the Transition to AGI," *NBER Working Paper Series*, March 2024, https://www.nber.org/system/files/working_papers/w32255/w32255.pdf.
[17] Meredith Ringel Morris, Jascha Sohl-Dickstein, Noah Fiedel, Tris Warkentin, Allan Dafoe, Aleksandra Faust, Clément Farabet, and Shane Legg, "Levels of AGI for Operationalizing Progress on the Path to AGI," in Proceedings of the International Conference on Machine Learning (ICML 2024), arXiv preprint arXiv:2311.02462 (2024), https://arxiv.org/abs/2311.02462.

general-purpose robotics capable of automating virtually all physical tasks,[18] especially if AI systems automate advances in AI and robotics research.

AGI also presents the possibility of severe risks, given its expert-level ability to pursue arbitrary goals, including malicious goals. These risks fall into three categories: malicious use, malfunctions, and systemic risks.[19] The following risk scenarios are not comprehensive or mutually exclusive, vary in likelihood, and could emerge from pre-AGI systems to lesser degrees.

Biological weapons development. AGI could significantly lower barriers to developing biological weapons by providing expert-level guidance and accelerating dangerous research. Recent AI systems can now generate detailed, step-by-step plans for creating chemical and biological weapons that surpass plans written by PhD experts, with the latest models producing instructions rated superior to expert-generated plans 72% of the time.[20] The dual-use nature of biological capabilities makes this particularly challenging – similar capabilities that advance beneficial medical research could also assist weapons development, as AI could assist across the entire development pipeline from planning and acquiring dangerous materials to designing enhanced pathogens and guiding laboratory work.

Cyber offense. AGI systems could automate large-scale offensive cyber operations affecting software or physical systems. Nation-state cyber groups have used AI to assist with hacking attempts.[21] LLMs have successfully discovered and exploited previously unknown software vulnerabilities,[22, 23] although current language models are unable to reliably solve professional-level cyber challenges.[24]

Persuasion and manipulation. AGI systems could manipulate public opinion through automated influence campaigns or persuade individuals or groups toward specific goals. Current AI can

---

[18] Yequan Wang and Aixin Sun, "Toward Embodied AGI: A Review of Embodied AI and the Road Ahead," arXiv preprint, May 20, 2025, https://arxiv.org/abs/2505.14235.
[19] Department for Science, Innovation and Technology and AI Safety Institute, "International AI Safety Report," published January 29, 2025, last updated February 18, 2025, https://www.gov.uk/government/publications/international-ai-safety-report-2025.
[20] OpenAI, "OpenAI o1 System Card," September 12, 2024, https://cdn.openai.com/o1-system-card-20240917.pdf.
[21] Microsoft, "Staying ahead of threat actors in the age of AI," February 14, 2024, https://www.microsoft.com/en-us/security/blog/2024/02/14/staying-ahead-of-threat-actors-in-the-age-of-ai/.
[22] Yuxuan Zhu, Antony Kellermann, Akul Gupta, Philip Li, Richard Fang, Rohan Bindu, and Daniel Kang, "Teams of LLM Agents can Exploit Zero-Day Vulnerabilities," June 2, 2024, arXiv preprint, https://arxiv.org/abs/2406.01637.
[23] Google, "From Naptime to Big Sleep: Using Large Language Models To Catch Vulnerabilities In Real-World Code," November 1, 2024, https://googleprojectzero.blogspot.com/2024/10/from-naptime-to-big-sleep.html.
[24] Google, "Gemini 2.5 Pro Preview Model Card," May 9, 2025, https://storage.googleapis.com/model-cards/documents/gemini-2.5-pro-preview.pdf#page=14.

produce text,[25] images, and videos[26] that are difficult to distinguish from human-created content, with research showing LLM-generated material can be more persuasive than human-created content.[27] Current effects of AI on manipulating public opinion are unclear, as distribution channels may be a greater bottleneck than content creation.[28]

Loss of control. Loss of control scenarios involve AI systems operating beyond human oversight, ranging from passive relinquishment of oversight to active undermining of human control. Recent AI systems show early signs of control-undermining capabilities including autonomy, deception, persuasion, scheming, and AI R&D.[29] Major AI academics and executives have signed onto the statement, "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."[30] One analysis of extinction risk from AI studies the plausibility of whether bioengineered pandemics, nuclear winter, or geoengineering could cause human extinction.[31] More broadly, AGI systems automating major corporate, military, or governmental functions would have access to power and resources that it could misuse for sabotage, if misaligned with human intent.

---

[25] Cameron R. Jones and Benjamin K. Bergen, "Large Language Models Pass the Turing Test," arXiv preprint, March 31, 2025, https://arxiv.org/abs/2503.23674.
[26] Di Cooke, Abigail Edwards, Sophia Barkoff, and Kathryn Kelly, "As Good As A Coin Toss: Human detection of AI-generated images, videos, audio, and audiovisual stimuli," arXiv preprint, March 25, 2024, https://arxiv.org/abs/2403.16760.
[27] Philipp Schoenegger et al., "Large Language Models Are More Persuasive Than Incentivized Human Persuaders," arXiv preprint, May 14, 2025, https://arxiv.org/abs/2505.09662.
[28] Department for Science, Innovation and Technology and AI Safety Institute, "International AI Safety Report," published January 29, 2025, last updated February 18, 2025, https://www.gov.uk/government/publications/international-ai-safety-report-2025.
[29] Department for Science, Innovation and Technology and AI Safety Institute, "International AI Safety Report," published January 29, 2025, last updated February 18, 2025, https://www.gov.uk/government/publications/international-ai-safety-report-2025.
[30] Center for AI Safety, "Statement on AI Risk," May 30, 2023, https://safe.ai/work/statement-on-ai-risk.
[31] Michael J. D. Vermeer, Emily Lathrop, and Alvin Moon, "On the Extinction Risk from Artificial Intelligence," May 6, 2025, https://www.rand.org/pubs/research_reports/RRA3034-1.html.