Lernende Systeme
GERMANY'S PLATFORM FOR ARTIFICIAL INTELLIGENCE

Artificial General Intelligence (AGI)

Between promises and reality

AT A GLANCE

Artificial General Intelligence (AGI)

- refers to AI systems that do not yet exist, which process information, learn and act flexibly, autonomously and across contexts – similar to humans.
- is currently defined in very different ways: from functional intelligence to conscious experience.
- is interpreted differently by different actors – for example, as a promise of progress, a risk or a science fiction scenario.
- is the subject of philosophical debates about consciousness, physicality and creativity: What distinguishes "real" intelligence?
- exemplifies the tension between technological feasibility, fundamental philosophical questions and social imagination.

The term "AGI" is used with very different meanings and goals – from technical specification to vision of the future. This ambiguity leads to exaggerated expectations and distorted public debates.

Context

The distinction between so-called weak and strong Artificial Intelligence (AI) was coined back in the 1980s. While weak AI performs specific tasks, strong AI is expected to reproduce or simulate human intelligence in its entirety, including consciousness and intentionality. The idea of formally capturing human cognition goes back even further: as early as 1943, Warren McCulloch and Walter Pitts proposed a simplified model of neural activity that served as the theoretical basis for artificial neural networks. However, it soon became clear that such models could only represent a very limited aspect of human intelligence. Today, the term "Artificial General Intelligence" (AGI) has increasingly replaced what was understood as strong AI. Despite enormous technological advances in recent years, today's AI systems remain far from AGI. AGI is associated with high hopes for medical, scientific and social breakthroughs, but also with profound questions about control, responsibility and the boundaries between humans and machines. It is therefore not only a technical goal, but also a reflection of social ideas, expectations and fears.

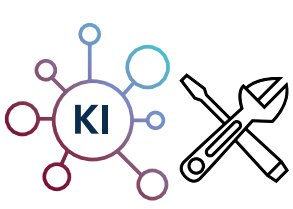
What is Artificial General Intelligence?

There is currently no generally accepted definition of AGI, nor is there any test that would confirm its achievement. Different research disciplines, companies and political actors have different priorities: some definitions emphasise cognitive flexibility and the ability to learn autonomously, while others focus on consciousness, independent goal setting or transfer performance. A central problem with all approaches is that they are mostly based on an idea of human intelligence – but even this is not clearly defined. In psychology, neuroscience and philosophy, there are competing approaches to what constitutes intelligence. This double ambiguity – in humans as well as in machines – makes it clear that AGI is not just a technical concept, but also a projection space for scientific, economic and social interests.

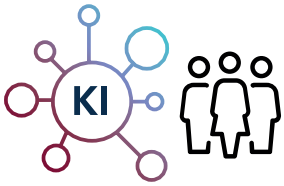
Actors	Characteristics of AGI	Interest	Criticism
Research	■ Cognitive flexibility ■ Autonomous learning ■ Transfer performance ■ (Partial) awareness	■ Theoretical understanding of intelligence ■ Disciplinary paradigms (e.g. psychology vs. computer science)	■ Inconsistent definition of (human) intelligence
Companies	■ Performance ■ Adaptability ■ Broad applicability	■ Strategic communication ■ Investment promotion ■ Market advantages through innovation narratives	■ Definitions mostly marketing-driven ■ Risk of overestimation/ ■ expectations
Science fiction	■ Superhuman intelligence ■ Emotional depth ■ Independent goals	■ Dramaturgy/cultural reflection ■ Visions of the future/warnings	■ Strong influence on social perceptions ■ Distortion of real research
Society	■ Human-like thinking and feeling ■ "All-rounder" ■ Consciousness ■ Autonomy	■ Orientation through media & fiction ■ Hopes for problem solving ■ Fears of losing control	■ Fiction instead of facts as a frame of reference ■ Polarised debate between hype and dystopia

Technical basics: How can AGI work?

The development of Artificial General Intelligence is one of the great challenges of modern AI research. The question being discussed is whether such a development is even possible and, if so, which approaches are promising.



Artificial Narrow Intelligence (ANI)
Can perform specific tasks



Artificial General Intelligence (AGI)
Can behave like humans in all tasks

Artificial Super Intelligence (ASI)
Smarter than humans –
just like in science fiction

Source: Plattform Lernende Systeme

Approaches to the development of AGI

The **scaling thesis**, currently espoused in particular by large tech companies, states that general intelligence could eventually be achieved through growth in model size, data volume and computing capacity. Proponents of this theory argue that today's AI models are already becoming increasingly powerful and capable of performing more complex tasks as they scale up – such as language models like GPT-5, which demonstrate better language comprehension as they grow in size.

Critics argue, however, that scaling alone will not be enough to achieve skills such as transfer learning, causal thinking, reasoning, decision-making, goal-setting or autonomy. Existing AI models still have difficulty transferring knowledge from one area to another or developing independent problem-solving strategies. The scaling thesis is also empirically questionable: the underlying correlations ("scaling laws") are not natural laws and seem to apply only within certain orders of magnitude. Accordingly, the performance of purely scale-based models cannot be increased indefinitely. Rather, there are indications that additional algorithmic innovations are necessary to achieve significant performance gains. Furthermore, it should be noted that current models already use the majority of high-quality training data and, accordingly, further scaling could fail due to a lack of data.

Actors	Characteristics of AGI	Interest	Criticism
Scaling	■ Ever larger models, more data and computing power → AGI emerges through sheer size	■ Rapid progress in certain tasks ■ Already powerful models today ■ Clear path to implementation	■ Lack of transfer, causality, reason ■ Resource-intensive ■ "Emergence" remains speculative ■ Perspective: lack of data
Hybrid AI	■ Combination of symbolic AI and neural networks	■ Better generalisation & explainability ■ Integration of world knowledge	■ Comprehensive integration still unresolved ■ Often less scalable than pure deep learning models ■ Still hardly standardised
New paradigms/ evolutionary approaches	■ AGI through new learning mechanisms, self-organisation & physicality	■ Orientation towards biological systems ■ Focus on motivation, goal setting, autonomy ■ More robust in the long term	■ Still in its early stages ■ Difficult to measure & complex

This is **where hybrid AI models** come in, combining data-driven and knowledge-based AI. They attempt to combine the strength of neural networks in handling large amounts of data with the logical structure of symbolic AI. The aim is to create systems that not only recognise patterns, but can also provide explanations and deal with explicit rules. Hybrid approaches are therefore considered promising for application areas where explainable and reliable behaviour is required, such as medicine or law. A successful example is AlphaFold, which predicts protein folding with previously unattainable accuracy – its developers were awarded the Nobel Prize in 2024.

In addition, there are **alternative paradigms** that seek to take completely different paths to AGI. These include **evolutionary approaches**, in which AI systems are developed over many generations through variation and selection – analogous to biological evolution. **Embodied AI** is also being discussed: here, it is assumed that true intelligence can only arise through physical interaction with the environment – i.e. through sensory perception, motor skills and embodied experience. These concepts tie in with developmental psychology and neurocognitive theories that emphasise the close connection between perception, body and thought.

Assessments by leading AI experts

The public debate on AGI is currently dominated by voices from the Anglo-American world. This reflects both the technological leadership of these countries and the prevailing narratives there about the future of AI.

AGI is achievable – soon

Sam Altman, CEO of OpenAI; Altman sees AGI as a realistic future within reach. He is focusing on scaling existing systems (transformer models) and investing heavily in hardware with OpenAI.

Gradual approach to AGI

Demis Hassabis, CEO of DeepMind, Nobel Prize winner; Hassabis takes a scientific, cautious approach. He sees AGI as a long-term goal, achievable through advances in reinforcement learning, memory architectures and simulation, among other things.

Concerns about loss of control

Geoffrey Hinton, AI pioneer, former Google employee, Nobel Prize winner; Hinton believes AGI is technically possible, but expresses concern about loss of control over powerful AI systems and warns against excessive euphoria.

General scepticism towards AGI

Yann LeCun, chief AI researcher at Meta, Turing Award winner; LeCun emphasises that current systems do not master fundamental skills such as causal reasoning or common sense and calls for new approaches beyond deep learning.

Call for hybrid approaches

Gary Marcus, AI critic, neuroscientist; Marcus considers deep learning alone to be insufficient and calls for a combination of symbolic logic and statistical methods to achieve what he believes to be true AGI.

Criticism of current AGI narratives

Margaret Mitchell, AI ethicist, formerly at Google AI Ethics; According to Mitchell, the AGI debate is often scientifically vague and obscures narrative power relations. She calls for more focus on fairness, responsibility and the real-world impacts of AI today.

Warning against AI with consciousness

Thomas Metzinger, philosophy professor, former member of an EU expert group on AI ethics; Metzinger suggests, as a thought experiment, that the development of AI systems with consciousness would lead to a massive increase in potentially suffering artificial subjects. He advocates a moratorium on development until ethical aspects have been sufficiently considered.

Criticism of power structures and lack of regulation

Meredith Whittaker, president of the Signal Foundation; Whittaker does not see AGI primarily as a technical issue, but rather a political issue. She warns against the concentration of AI systems in the hands of a few tech companies and calls for democratic control, transparency and protection of public infrastructure.

Philosophical and social classification

The debate surrounding Artificial General Intelligence raises not only technical questions, but also philosophical and social ones. A key issue here is the **definition of intelligence** itself: is the ability to solve problems and learn sufficient, or are creativity, consciousness, physicality and autonomy also inseparable parts of it?

The debate surrounding AGI inevitably pushes us to the limits of our understanding of intelligence, subjectivity and creativity – and ultimately reflects questions we ask ourselves about our own nature. Whether machines will ever be comparable to humans in terms of consciousness, creativity or autonomy (or whether they can only simulate these qualities) remains an open question for now. But even the discussion of these possibilities shows how much the AGI concept is not only a technical goal, but also a cultural mirror. What we understand by "general intelligence" – and whether we want to recognise it in machines – is closely linked to our ideas about what it means to be human.

	Human	AGI
Consciousness 	■ Intelligence is inextricably linked to consciousness (subjective experience, perception, self-awareness)	■ Functions such as memory and language processing can be explained technically, but not the subjective experience ("hard problem of consciousness"). or ■ Functionalist AI research: AGI with limited cognitive functions is possible without true consciousness.
Physicality & autonomy 	■ Humans experience and understand the world not only through abstract symbols, but through their bodies, senses, emotions and social interactions ■ Autonomy is based on biological drives: survival, pain avoidance, pleasure, reproduction	■ No sensory experience, no biological drives – and therefore no natural motivation or ■ Digital systems could also develop forms of goal-setting, for example through reward mechanisms
Creativity 	■ Creative intention: New things are created with meaning, contextual understanding and cultural embeddedness ■ Combinatorial, exploratory and transformative creativity: People can create completely new things on their own initiative	■ Creativity based on statistical combinations, not on inner intention or cultural embeddedness ■ It is questionable whether AI will ever be capable of transformative creativity (fundamental rethinking of cultural conventions)

Legal and ethical challenges

The development of Artificial General Intelligence goes far beyond a technical vision. It touches on fundamental questions of law, ethics and social coexistence.

Liability: Who is responsible for AGI decisions?

Who is responsible if, in a hypothetical future, an AGI decision leads to damages: The developers? The operating company? The users? The AGI itself? The more autonomously a system acts, the more difficult it becomes to assign clear responsibilities. Regulations such as the GDPR already address automated decisions and impose strict requirements for transparency and accountability. However, they presuppose that there is a comprehensible human chain of responsibility. In a true AGI, which by definition makes independent decisions, these prerequisites would no longer apply – and with them the framework for applying existing legal principles.

Rights for machines? The debate on the legal and ethical status of AGI

With increasing autonomy, the question arises as to whether an AGI with consciousness, its own goals, or a kind of experience can continue to be regarded as a mere tool – or whether it should be granted its own legal status, analogous to legal entities such as companies. Some ethicists and lawyers argue that moral status could also give rise to rights – such as the right not to be switched off or to have control over oneself.

If an AGI were recognised as the bearer of moral claims, this could result not only in protective rights but also in duties of care – similar to those of natural persons. This raises further questions: Would it be permissible to force an AGI to work? Could cloning bans or protective mechanisms analogous to those in bioethics be applied? Recognised moral or legal status could also give rise to obligations: just as companies are liable for certain types of damage, it could be argued that AGI systems could also be assigned normative responsibility – for example, through digital trust systems or vicarious liability.

Regulation: Is existing AI law sufficient?

Current laws – such as the European Union's AI Act – are primarily geared towards the risks posed by today's AI systems. With specific requirements for "general purpose AI" (GPAI), the AI Act does contain initial regulations for broadly applicable AI models. However, these are unlikely to be sufficient to cover hypothetical AGI systems that pursue their own goals or act in social contexts. This raises the question of whether adjustments to the AI Act or even a separate legal regime for AGI are needed – comparable to international law or special bioethical regulations.

Where is AI heading?

Those who believe that AGI will be achievable in the near future often also believe that its further development will lead to superintelligence that will surpass human intelligence cognitively. This is often accompanied by fears that superintelligence could seek to dominate humanity. In the short term, however, it is not the machine itself that is likely to pose a threat, but rather the question of who controls it: Could individual actors – such as large technology companies or states – use superintelligence to expand their power?

Expertise from Plattform Lernende Systeme

For AI applications, everything is the same: they work just as well for an intelligent traffic light system as they do for analysing love poems. AI does not understand anything, but calculates patterns. Its view of the future is a probability forecast. It cannot hope for anything because it has nothing to lose. AI lacks the experience of physical existence. Even though AI language models can copy everything humans say about consciousness, desires or solidarity, they connect with it as much as they do with the traffic light system mentioned above. Superintelligence is a hypothetical concept, an interesting thought experiment, but nothing more.

Jessica Heesen, Professor at the International Centre for Ethics in Science and Technology (IZEW) at the University of Tübingen

AGI remains a vague concept. Despite the successes of deep learning, AI systems lack common sense: they do not think logically, fail when faced with new situations and consume vast amounts of resources. Pure scaling will not lead us to AGI. Instead, we need to combine AI with cognitive science and unite deep learning with symbolic reasoning. This will enable us to create intelligent AI that solves complex problems, builds trust and efficiently handles citizen concerns in public administration, for example.

Kristian Kersting, Professor of Machine Learning and Artificial Intelligence at TU Darmstadt and Co-Director of hessian.AI

AI is more than an ordinary tool and less than a human subject. The probability of a superintelligence that wants to destroy humanity is therefore low. The dangers lie in the enormous scope for power that AI opens up for a few, as well as in the lack of predictability and control that is a structural feature of AI. In addition, AI is increasingly taking over tasks that humans are actually good at and enjoy doing, such as interpersonal or creative work. I would like to see more development in problem areas where we do not yet have good solutions.

Catrin Misselhorn, Professor of Philosophy at Georg August University of Göttingen

Outstanding issues

- **Security:** How can we ensure that AGI systems act in accordance with human values – even with open learning behaviour and growing autonomy?
- **Control:** How can we prevent AGI from becoming autonomous or being misused by a small group?
- **Philosophy:** What characteristics (e.g. consciousness, understanding, physicality) are necessary for something to be considered "true" intelligence?
- **Governance:** Are national regulations sufficient, or are international rules and institutions needed for AGI development?
- **Influencing factors:** How strongly do economic interests, media narratives and science fiction shape our expectations of AGI – and how does this affect research and politics?

Further reading

Altman, S. (2023). Sam Altman on AGI: Scaling large language models is not enough. The Decoder. https://the-decoder.com/sam-altman-on-agi-scaling-large-language-models-is-not-enough/

Bender, E. M., Gebru, T. et al. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT). https://arxiv.org/pdf/2002.06177

Hassabis, D. (2025). Demis Hassabis: "Why AI must be built responsibly". TIME 100 Interview. https://time.com/7277608/demis-hassabis-interview-time100-2025/

Krasheninnikov, A., Yao, J., & Sutskever, I. (2023). On the measurement of AGI progress. arXiv. https://arxiv.org/pdf/2311.02462

Mitchell, M., Wu, S., Zaldivar, A. et al. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*). https://arxiv.org/pdf/1810.03993

University of Toronto (2023). Risks of artificial intelligence must be considered as technology evolves: Geoffrey Hinton. https://web.cs.toronto.edu/news-events/news/risks-artificial-intelligence-must-be-considered-technology-evolves-geoffrey-hinton

The content of **AI** AT A GLANCE may be used for editorial purposes, provided that the source Plattform Lernende Systeme is cited.

Imprint
Expertise: Jessica Heesen, Kristian Kersting, Catrin Misselhorn
Editors: Paul Grünke, Birgit Obermeier
Publisher: Lernende Systeme – The Platform for Artificial Intelligence | Managing Office | c/o acatech | Karolinenplatz 4 | D-80333 Munich
presse@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de
Status: September 2025 | Image credits: Thilo Schoch, University of Darmstadt, Stefanie Trenz | Follow us on LinkedIn and YouTube.
Translation with help of DeepL

With funding from the:

Federal Ministry of Research, Technology and Space

acatech NATIONAL ACADEMY OF SCIENCE AND ENGINEERING