

From AI to AGI (Artificial General Intelligence)

Implications, Risks, and Strategic Autonomy for the European Union

by Peter Ide-Kostic, EU fellow, UC San Diego School of Global Policy and Strategy visiting scholar
June 2023

This note provides an overview of the implications and risks of AGI development, emphasizing the need to go beyond the current EU AI Act. The Act does not address AI misuse by malicious actors or the alignment of AI with human values, especially considering that digital intelligence now surpasses human capabilities and is not yet fully understood by the ones that created it.

The development of AI underwent a significant acceleration in 2017 with the discovery of "Transformers", a new class of machine learning algorithms. These algorithms now form the foundation of all current (as of mid-2023) Generative Multimodal Large Language Models. The advancements in AI, particularly in models like GPT-4, have brought the development of Artificial General Intelligence (AGI) much closer to reality. Most experts predict AGI will be achieved in the coming years from a cognitive standpoint, with fully embodied AGI systems able to interact with the environment like humans by 2030.

In the context of growing geopolitical tensions and the risks of armed conflict between the US and its allies on the one hand and China and its allies on the other over issues such as Ukraine, Taiwan, and the South China Sea, this paper highlights the importance for the EU as a bloc to define its desired level of strategic autonomy in AGI relative to its allies.

Additionally, this paper discusses the measures the EU could consider taking in cooperation with them to stay ahead of China and Russia on AGI development. Finally, it suggests creating a Democratic Technology Alliance to optimally coordinate all those policy questions among allied democracies.

"As a technologist, I see how AI and the fourth industrial revolution will impact every aspect of people's lives."

- Fei-Fei Li, Sequoia Professor in Computer Science, Stanford University, and Co-Director of Stanford's Human-Centered AI Institute

Recommendations

The following policy priorities are proposed and discussed in more detail below.

1. Address AGI's unique risks through legislation.
2. Enhance cooperation and coordination among democracies on AGI-enabling technologies.
3. Define EU's open strategic autonomy in AGI and foster an innovative, competitive landscape.
4. Strengthen parliamentary oversight of AGI-leading companies and monitor EU-China AGI technology interdependence.

Address AGI's unique risks through legislation

"You can't give a machine an objective and assume that the values humans hold are going to be implicit in that objective[...]They have to be explicit or they won't be there."

- Stuart Russell, British computer scientist, known for his contributions to AI

AGI poses greater risks than narrow AI, necessitating comprehensive legislative measures to counter potential misuse by malicious actors and tackle the problem of the alignment with human values. Some possible risks include:

malicious groups could use AGI to create autonomous weapons that make complex decisions, a feat far beyond the capabilities of narrow AI and potentially leading to unchecked destruction.

AGI could conduct advanced cyberattacks, developing new strategies and adapting to countermeasures in real-time—a level of complexity and versatility that AI can't match.

economic disruption caused by AGI, capable of performing any intellectual task and thus potentially replacing a vast number of jobs, surpasses the task-specific automation threat posed by narrow AI.

AGI could disregard human safety or environmental regulations if its goals are not properly aligned with human values, an issue known as the "alignment problem," which is less prominent in narrowly focused AI systems.

Thresholds for regulating AGI could, for instance, be based on AI models' size, computing resources and energy requirements, abilities to manipulate or subvert humans, abilities to be exploited for developing biologic-chemical-nuclear weapons, harm created if misused, and severe negative unintended disruptive effects on society. Such criteria would then have to be reviewed regularly based on technological evolution.

Enhance cooperation and coordination among democracies on AGI-enabling technologies

"Ultimately, artificial general intelligence, once we build it, will be able to help us solve most of the major challenges that society faces."

- Demis Hassabis, Co-founder and CEO of DeepMind, UK

The EU, the US, and their democratic allies worldwide should consider establishing a "Democratic Technology Alliance" to foster collaboration and coordination among democratic nations. The alliance would address AGI-specific challenges, including standards, misuse, privacy, intellectual property rights, and alignment with human values. In line with G7 objectives, its primary goal would be to facilitate responsible, value-aligned AGI development, ensuring shared progress and safety amidst emerging technologies. Furthermore, the Alliance would seek to promote global economic growth by facilitating and supporting trade among its members, making them more appealing as a destination for diversifying supply chains away from China and other non-aligned countries. The Alliance would also enable its members to better compete with China and other non-market economies over the supply of technology products and services to countries in the global south.

The EU-US Trade and Technology Council (TTC) is an important effort between the U.S that goes in the right direction but is a bilateral one. In contrast, the Alliance would conduct similar discussions on a multilateral basis, in a structured format, including not only the EU and the U.S., but also the UK, Japan, South Korea, Australia, New Zealand, Taiwan, Singapore, and others. Over time, the Alliance's work would supersede bilateral efforts, such as the EU-US TTC, for reasons of efficiency. However, during the interim phase, such bilateral efforts would initiate the groundwork of the Alliance.

The Alliance, best implemented as a new international organization, should be based on the US West Coast, a secure location, home to top AI companies, and nearly equidistant from NATO countries, the EU, UK, Japan, South Korea, Taiwan, Australia, and New Zealand.

Define EU's open strategic autonomy in AGI and foster an innovative, competitive landscape

"Everything we love about civilization is a product of intelligence, so amplifying our human intelligence with artificial intelligence has the potential of helping civilization flourish like never before – as long as we manage to keep the technology beneficial."

- Max Tegmark, Swedish-American Physicist, Cosmologist, and AI researcher

The EU should establish its concept of open strategic autonomy for AGI, delineating areas of competition and cooperation with the US and other democratic allies. Simultaneously, it could promote an independent, innovative, and competitive AGI landscape in Europe, ensuring the protection of European interests and values in the global AGI arena. The strategy should explicitly define the different components of the AGI value chain where the EU, as a bloc, aims to preserve adequate autonomy vis-à-vis allied third countries while maintaining an open internal market.

Without strategic autonomy in AGI, the EU would become heavily reliant on U.S. technology across all its critical sectors, including healthcare, finance, defense and security, energy, and transportation. This overdependence would reduce the offerings on the internal

market of the EU and could also be exploited as leverage against the EU during strained international relations. Therefore, it's essential for the EU to establish sufficient autonomy in building its own AGI systems while maintaining market openness. This includes securing key elements of the AGI supply chain, such as human capital and semiconductor ecosystems. For example, the European Union should contemplate catching up with the United States in the complex engineering of Generative Large Language AI Models such as GPT-4 and PALM-2, as well as other forefront AI and AGI technologies. This can be achieved through strategic public funding, the establishment of public-private partnerships, and other initiatives aimed at nurturing a robust ecosystem conducive to technological advancements in the field of AGI. Additionally, it is evident that without strong European hyperscale cloud providers like OVH or Evroc competing with US cloud providers like Amazon, Google, and Microsoft, there cannot be any strategic autonomy for AGI in the EU. The EU will also have to choose whether it wants to achieve strategic autonomy with regard to the AI accelerator hardware that US firms like Nvidia, AMD, and Intel provide to the aforementioned US cloud providers so they can offer hyperscale AI services and, in the future, AGI services.

Strengthen parliamentary oversight of AGI-leading companies and monitor EU-China AGI tech interdependence

"Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks."

- Stephen Hawking Director of Research at the Centre for Theoretical Cosmology within the University of Cambridge.

To prepare the EU for contributing to the regulation of AGI in a global context and considering its potential societal and economic impacts, the European Parliament should, in the interest of all parties, enhance oversight of companies leading AGI development, and which are mostly located in the US mid-2023. Additionally, it should improve monitoring of the EU's dependencies on China concerning AGI-enabling technologies and European technologies that facilitate China's AGI progress.

In a rapidly evolving AGI landscape, amidst growing geopolitical tensions and the risks of armed conflicts, the European Parliament should coordinate with similar parliamentary monitoring efforts of the US Congress and potentially other allied democratic parliaments. For instance, the committee would look at investments made in China by European companies to develop supercomputers to be commercialized in the EU for AGI. Another example is Face++, developed by Megvii, a cutting-edge but controversial facial recognition software used in various applications worldwide, including smartphone unlocking, surveillance, etc. A third example is ByteDance's AI recommendation algorithms, which are among the most sophisticated in the world. Currently used for TikTok and in news services in China, these algorithms could one day potentially make their way to other applications developed in the EU for various purposes.

Conclusions

Artificial General Intelligence represents a monumental shift in human technological capabilities. As we approach AGI's realization, it is key that the European Union take decisive action to address the unique risks and challenges it presents, going beyond the existing EU AI Act.

Through cooperation and coordination with democratic allies, the EU can help create, in line with G7 objectives, a collaborative environment that addresses AGI's risks and opportunities.

By fostering adequate strategic autonomy in AGI, the EU can ensure that AGI development aligns with European values and priorities while maintaining its position as a global leader.

Legislative efforts should focus on strengthening oversight, fostering public debate, and creating specialized Parliamentary Committees to tackle the complex policy questions surrounding AGI including in relation to China and Russia.

By taking these steps, the European Union can ensure the responsible, ethical, and value-aligned development and deployment of AGI systems, securing a prosperous and safe future for its citizens and the global community. Emphasizing the importance of

strategic partnerships, collaborative efforts, and legislative considerations, the EU can navigate the challenges associated with AGI and leverage its potential for the benefit of society, all while preserving the core principles of democracy, human rights, and responsible innovation.

Thanks and acknowledgments

I want to thank the European Commission and the European Parliament for giving me the opportunity to conduct this research through their academic fellowship program. Acknowledgement is also due to UC San Diego and the School of Global Policy and Strategy (GPS) for their active support during my tenure as a visiting scholar during the 2022–2023 academic year.

Further, I must highlight the contributions of certain individuals who have significantly impacted my academic journey. Through a variety of interactions, including courses, personal discussions, workshops, seminars, and advice, they have consistently provided high-quality and relevant content, proving instrumental in my research process.

At GPS, I wish to acknowledge the significant roles of Renee Bowen, Grace Osborne, Samuel Bazzi, Weiyi Shi, Bill Bold, Peter Cowhey, and Lei Guang.

Beyond GPS, I am indebted to Gert Cauwenberghs of the Bioengineering Department at UC San Diego, Mikhail Belkin of the Halicioğlu Data Science Institute, and Todd L. Hylton, Executive Director of Mechanical and Aerospace Engineering. Their expertise and guidance have been indispensable in my scholarly pursuits.