A Proposal for a Definition of General Purpose Artificial Intelligence Systems

Carlos I. Gutierrez, Anthony Aguirre, Risto Uuk, Claire C. Boine, and Matija Franklin

Abstract— **The European Union (EU) is in the middle of comprehensively regulating artificial intelligence (AI) through an effort known as the AI Act. Within the vast spectrum of issues under the Act's aegis, the treatment of technologies classified as general purpose AI systems (GPAIS) merits special consideration. Particularly, existing proposals to define GPAIS do not provide sufficient guidance to distinguish these systems from those designed to perform specific tasks, denominated as fixed-purpose. Thus, our working paper has three objectives. First, to highlight the variance and ambiguity in the interpretation of GPAIS in the literature. Second, to examine the dimensions of generality of purpose available to define GPAIS. Lastly, it proposes a functional definition of the term that facilitates its governance within the EU. Our intention with this piece is to spark a discussion that improves the hard and soft law efforts to mitigate these systems' risks and protect the well-being and future of constituencies in the EU and globally.**

Index Terms— **Artificial intelligence, European Union, governance of technology, technology risk management**

I. INTRODUCTION

IN 2021, the European Union (EU) published its first draft of the Artificial Intelligence (AI) Act. Among its proposals, this comprehensive regulation creates a risk classification framework that groups AI systems into four categories: unacceptable, high risk, limited risk, and low risk. Acknowledging the changing nature of technology, authorities incorporated a provision to continuously assess the risk classification of systems. In making these determinations, the EU is instructed to consider "the intended purpose of the AI system."[1] This provision raises a critical issue, which is that AI systems may escape or evade the Act's safeguards because there can be a complex mapping between who develops and deploys them, the tasks they perform, and the purpose(s) they serve as a product. Hence, discussion of technologies lacking an intended purpose has crystallized as a debate around the meaning of the term general purpose AI system (GPAIS).

As it stands today, no guidelines explain the inclusion criteria for AI systems that classify as GPAIS. Within the AI Act context, the existing definition has many opportunities for improvement. The Slovenian EU presidency defined GPAIS as an "AI system… able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation, etc."[2] The French EU presidency further emphasizes that GPAIS: "may be used in a plurality of contexts and be integrated in a plurality of other AI systems."[3] Outside the EU context, the term GPAIS is seldomly and haphazardly used to describe AI systems that vary considerably in terms of autonomy, agency, modality, and training methods.

How the discussion unfolds on the definition of GPAIS is crucial to the governance of AI in the EU and as a signal for global policymaking. To complement the debate, this piece is written with three objectives in mind. First, to highlight the variance and ambiguity in the interpretation of GPAIS in the literature. Second, to examine the dimensions of generality of purpose available to define GPAIS. Lastly, to propose a functional definition that facilitates this technology's governance. Our intention with this piece is to spark a discussion that improves the hard and soft law efforts to mitigate these systems' risks and protect the well-being and future of constituencies in the EU and globally.

II. GPAIS: A DEFINITIONAL MORASS

The AI Act has generated a need for an actionable definition of GPAIS where none currently exists.[4]–[9] Prior to its adoption by the EU, scant literature identifies AI systems as GPAIS. When it does, it describes a range of technologies with vastly different levels of competency.

On one end of the spectrum, GPAIS possess skills comparable to or greater than humans. Scholars have characterized these technologies as having enough self-sufficiency to adapt to unknown environments, make decisions with limited resources, and function in complex domains that are currently reserved for humans due to the need for contextualization.[10]–[14] This conception of GPAIS resembles the original notion of the General Problem Solver (or the oft-employed but also imprecisely-defined Artificial General Intelligence), whereby human aptitudes such as creativity, dexterity, and ingenuity manifest in a machine system.[15]

On the other end of the spectrum, AI systems with limited functionality are occasionally referred to as GPAIS.[16] For instance, a medical application in the form of a disease-agnostic simulation able to predict "the consequences of various treatment or policy choices" is described as a GPAIS. Similarly, others envision it as a technology that engages with games of different rules, a language model that generates strings of text, or a text-processing tool that analyzes patent data to assist government examiners with a variety of duties.[17]–[19]

III. PERSPECTIVES ON GENERALITY OF PURPOSE

Identifying the technologies that fit under the GPAIS umbrella requires agreement on the meaning of the acronym's first two initials. We begin by noting that there are several dimensions in which generality of purpose can be construed, corresponding to different options for the EU to consider as most appropriate for their constituency and regulatory objectives. In this section, we

focus on four alternatives: ability, domain, task, and output.[20] Importantly, we also emphasize that the *generality* of a system is distinct from its *capability*. Capability points to a system's competence, accuracy, or effectiveness; generality expresses how broadly and evenly capability is distributed

Generality of purpose could center on a system's abilities, otherwise described as the category of actions or processes it can effect. The literature has identified several fields to organize abilities such as language, vision, robotics, interaction, understanding, reasoning, and search.[5], [21] Some existing systems have one core ability (such as image recognition); others (e.g., Dall-E and Open Diffusion) include combinations – such as text processing plus image generation – and could be considered more general in this dimension.

Domain (also known as application or economic context space in the literature) is another option for contextualizing general purpose.[18], [21] It refers to the sector(s) of the economy where a technology provides its solutions (e.g., healthcare, education, defense, infrastructure, etc.). This concept is a proxy for a system's competence in fulfilling the needs of stakeholders in a particular sector. If we take education as an example, this entails satisfying, to a degree, the needs of teachers, parents, administrators, students, among others. AI systems servicing multiple domains would be thought of as more general than those that serve one or a portion of a domain.

The specification of a distinct type of problem, the definition of goals, or actions a system must achieve, is called a task.[5], [18], [22] Tasks relate to, but are distinct from abilities. Abilities represent the overarching range of actions, goals, or problems that a system performs, while the term task is used to denote a particular concrete combination of these elements. For example, most large language models have the ability to process and produce text. But "processing text" is not a well-specified task. Rather, tasks would include summarizing text, labeling data, answering classes of questions, etc. An AI system that is more general in this dimension would perform a wider variety of distinct tasks.

Lastly, there is output. This term is the final chain in the AI process and identifies the results of a task.[18], [23] This can be exemplified by a unique essay or image emanating from a single sentence prompt, the set of actions taken in playing a game or driving a car, or validation of an individual's identity via facial recognition. An AI system's generality within this dimension can be gauged by the variety of different outputs that are produced.

These distinct dimensions of generality would classify AI systems in different ways. For example, an image classification system would be task-narrow, ability-narrow, and probably output-narrow, but could be *domain*-general, by being incorporated in diverse applications. Meanwhile, a complex *task* (like autonomous driving) could require many diverse *abilities*. In terms of capability versus generality, one could have a highly capable but narrow system like AlphaGo, or a very general system with low levels of capability, making it mediocre in the many things it does.

IV. PROPOSING A DEFINITION FOR GPAIS

Our approach in discerning GPAIS is to propose a task-based definition. In essence, a task is the building block of what AI systems achieve. Tasks can be created to focus on one ability

and domain or permeate through several. In addition, regardless of how unique an AI system's outputs are, its tasks remain a factor that can consistently be used to measure and compare the activities of different systems.

We divide AI systems into two groups. First there are fixed-purpose systems. As their name suggests, they are created with a specific objective and can only accomplish tasks they are trained to perform. For instance, a fixed-purpose translation AI system is limited to translating text. Although systems may be coupled or combined with others to complete a higher number of tasks, if each of them is limited to performing the task they are originally trained for, they qualify as fixed-purpose.

GPAIS can also be trained to complete specific tasks. However, these systems differ from their fixed-purpose counterparts because they can perform tasks that they were not originally trained for. This is due to a combination of factors such as the quantity of input data or model structure. For instance, even though GPT-3 was trained to predict the next word in a string of text, it has been fine-tuned to support new tasks such as translation or answering questions. Considering the factors above, we propose the following definition of GPAIS:

An AI system that can accomplish or be adapted to accomplish a range of distinct tasks, including some for which it was not intentionally and specifically trained.

This definition highlights the distinctive feature of a GPAIS, which is its ability to complete tasks outside of those it is specifically trained for. Thus, it includes unimodal (e.g., GPT-3 and BLOOM) and multimodal (e.g., Stable Diffusion and Dall-E) systems. It contains systems at different points of the autonomy and agency spectra, with and without humans in-the-loop. GPAIS also differ in terms of their learning potential. While GPT-3 does not retain information from each interactive session, MuZero starts with complete ignorance about its virtual surroundings and then continuously learns about them. The term GPAIS also encompasses systems with "emergent" abilities, meaning new and surprising abilities that manifest at threshold levels of model parameter count and/or training computation.[24] Lastly, GPAIS are trained through different methods. Whereas Gato uses supervised learning, MuZero is based on reinforcement learning.[25], [26] Intuitively, the proposed definition can be read as: in fixed-purpose AI systems we choose a set of tasks, then train a system to do those particular tasks. In a GPAIS, we train the system, then choose tasks for it to do (perhaps with additional tailoring or training).

Our proposed definition for GPAIS excludes a variety of systems that are designed and trained to complete concrete tasks and objectives. Among these are: image classifiers, voice production, and recognition systems, etc. We consider such technologies fixed-purpose systems that can be *incorporated* into a variety of other systems (and/or used in different domains), but are not themselves GPAIS based on our definition. Likewise, a product such as a voice assistant that incorporates a number of fixed-purpose technologies would be excluded from the definition as long as each task in its repertoire corresponds to a model trained specifically for a task

and it cannot perform distinct tasks beyond that.

As technology continues to progress, AI systems will jointly evolve with it. Efforts in research and development that began in the 1950's led to the creation of fixed-purpose systems. Now they are followed by GPAIS and in the future we can expect advances in capability and generality that supersede any current form of technology. Regardless of what name these systems are given, general or strong AI, our proposed definition is agnostic to these breakthroughs and includes them as a subset.

Our definition is largely qualitative, though one can consider a system to be "more" of a GPAIS to the degree that it performs a wider variety of tasks, less of which it was specifically and originally trained for. For some purposes it could be useful to consider *quantitative* measures that evaluate the degree of a system's generality and capability by assessing its performance according to evaluation metrics across a wide range of tasks.[2,26] The completion of such work requires significant further research.

V. IMPLICATIONS OF DEFINING GPAIS

No consensus definition of GPAIS is available and this has important repercussions on the governance of AI. As society attempts to promote safe and trustworthy AI systems, it needs clear guidance for discriminating between fixed-purpose and GPAIS, and for identifying their potential beneficial and harmful uses.

The EU is a pioneer in AI law. While the AI Act is a groundbreaking piece of proposed legislation, one of its weaknesses is its focus on the purpose of AI systems.[27] This approach works effectively with fixed-purpose systems because they are designed with a defined set of tasks in mind, making it possible for policymakers to pinpoint the border between a low and high risk system.

Conversely, GPAIS do not necessarily have an intended purpose. They can engage in tasks that are outside the scope of their initial development. This means that delimiting a risk profile via the sector-approach envisioned by Annex III of the AI Act draft is less effective. Moreover, their ability to function across market sectors means that a flaw in GPAIS may even have a systemic negative effect on the economy. This reality affects all parties (consumers, developers, and downstream firms that incorporate a technology into their offering, among others) that engage with and push the limits of AI systems. Therefore, EU regulators should pursue a regulatory framework that ensures the safety of GPAIS design, development, and deployment.

The definition proposed in this piece represents a first step. Through it, policymakers within and outside the EU are better equipped to classify systems placed within their market. The changing nature of technology will merit continuous updates to our understanding of GPAIS. In fact, members of the European Parliament have proposed the creation of a sub-group in the Artificial Intelligence Board to deal with questions of this nature.[28] Overall, we hope that this piece serves as a contributing factor to improve the predictability of AI governance. Lack of action in addressing the regulatory gap that can surface due to the EU's emphasis on a system's "intended purpose" may catalyze important long-term risks that the region and rest of the world should proactively avoid.

REFERENCES

[1] Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. 2021. Accessed: Apr. 27, 2022. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[2] Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Presidency compromise text. 2021. Accessed: Apr. 27, 2022. [Online]. Available: https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf

[3] Council of the European Union, Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union- Text de compromis de la présidence - Article 3, paragraphe 1 ter, Articles 4 bis à 4 quater, Annexe VI (3) et (4), considérant 12 bis bis. 2022. [Online]. Available: https://artificialintelligenceact.eu/wp-content/uploads/2022/05/AIA-FRA-Art-34-13-May.pdf

[4] S. Russell, "The history and future of AI," Oxford Review of Economic Policy, vol. 37, no. 3, pp. 509–520, 2021.

[5] J. Hernández-Orallo, "Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement," Artificial Intelligence Review, vol. 48, no. 3, pp. 397–447, 2017.

[6] N. Kagiyama, S. Shrestha, P. D. Farjo, and P. P. Sengupta, "Artificial intelligence: practical primer for clinical research in cardiovascular disease," Journal of the American Heart Association, vol. 8, no. 17, p. e012788, 2019.

[7] O. Strelkova, "Three types of artificial intelligence," 2017.

[8] E. A. Feigenbaum, "Artificial intelligence research," IEEE Transactions on Information Theory, vol. 9, no. 4, pp. 248–253, 1963.

[9] N. J. Nilsson, "Artificial intelligence prepares for 2001," AI Magazine, vol. 4, no. 4, pp. 7–7, 1983.

[10] P. Wang, "Toward a Unified Artificial Intelligence.," in AAAI Technical Report (1), 2004, pp. 83-.

[11] J. Bieger, K. R. Thórisson, B. R. Steunebrink, T. Thorarensen, and J. S. Sigurdardottir, "Evaluation of general-purpose artificial intelligence: why, what & how," Evaluating General-Purpose AI, 2016.

[12] N. J. Nilsson, "Human-level artificial intelligence? Be serious!," AI magazine, vol. 26, no. 4, pp. 68–68, 2005.

[13] J. P. Meltzer, "The impact of artificial intelligence on international trade," Brookings Institute, 2018.

[14] C. Szegedy, "A promising path towards autoformalization and general artificial intelligence," in International Conference on Intelligent Computer Mathematics, 2020, pp. 3–20.

[15] A. Newell, J. C. Shaw, and H. A. Simon, "Report on a general problem solving program," in IFIP congress, 1959, vol. 256, p. 64.

[16] C. C. Bennett and K. Hauser, "Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach," Artificial intelligence in medicine, vol. 57, no. 1, pp. 9–19, 2013.

[17] J. Somers, "How the artificial-intelligence program AlphaZero mastered its games," The New Yorker, vol. 3, 2018.

[18] OECD, "OECD Framework for the classification of AI systems." 2022. [Online]. Available: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1658796936&id=id&accname=guest&checksum=ACF491EF5BF19D52416E1D95710F9F05

[19] D. Alderucci and D. Sicker, "Applying artificial intelligence to the patent system," Technology & Innovation, vol. 20, no. 4, pp. 415–425, 2019.

[20] J. Hernández-Orallo, B. S. Loe, L. Cheke, F. Martínez-Plumed, and S. Ó hÉigeartaigh, "General intelligence disentangled via a generality metric for natural and artificial intelligence," Scientific reports, vol. 11, no. 1, pp. 1–16, 2021.

[21] R. Bommasani et al., "On the opportunities and risks of foundation models," arXiv preprint arXiv:2108.07258, 2021.

[22] K. R. Thórisson, J. Bieger, T. Thorarensen, J. S. Sigurðardóttir, and B. R. Steunebrink, "Why artificial intelligence needs a task theory," 2016, pp. 118–128.

[23] D. Ganguli et al., "Predictability and surprise in large generative models," 2022, pp. 1747–1764.

[24] J. Wei et al., "Emergent abilities of large language models," arXiv preprint arXiv:2206.07682, 2022.

[25] S. Reed et al., "A generalist agent," arXiv preprint arXiv:2205.06175, 2022.

[26] J. Schrittwieser et al., "Mastering atari, go, chess and shogi by planning with a learned model," Nature, vol. 588, no. 7839, pp. 604–609, 2020.

[27] C. Boine, "L'IA générale et la proposition de règlement de la Commission européenne," Dalloz IP/IT, vol. 2022, no. 2, Jan. 2022, [Online]. Available: https://www.dalloz-revues.fr/revues/Dalloz_IP_IT-750.htm

[28] Brando Benifei and Dragoş Tudorache, "Amendments 2356 - 2726: Draft report Brando Benifei, Dragoş Tudorache (PE731.563v01-00) Harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts." Jun. 13, 2022. [Online]. Available: https://artificialintelligenceact.eu/wp-content/uploads/2022/06/AIA-IMCO-LIBE-Report-All-Amendments-14-June.pdf