Article 51

Classification of general-purpose AI models as general-purpose AI models with systemic risk

Commentary by Gregor Gindlin | Submitted: January 2026

AI Act provision

Article 51

1. A general-purpose AI model shall be classified as a general-purpose AI model with systemic risk if it meets any of the following conditions:

(a) it has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks;

(b) based on a decision of the Commission, *ex officio* or following a qualified alert from the scientific panel, it has capabilities or an impact equivalent to those set out in point (a) having regard to the criteria set out in Annex XIII.

2. A general-purpose AI model shall be presumed to have high impact capabilities pursuant to paragraph 1, point (a), when the cumulative amount of computation used for its training measured in floating point operations is greater than 10^{25}.

3. The Commission shall adopt delegated acts in accordance with Article 97 to amend the thresholds listed in paragraphs 1 and 2 of this Article, as well as to supplement benchmarks and indicators in light of evolving technological developments, such as algorithmic improvements or increased hardware efficiency, when necessary, for these thresholds to reflect the state of the art.

Annex XIII: Criteria for the designation of general-purpose AI models with systemic risk referred to in Article 51

For the purpose of determining that a general-purpose AI model has capabilities or an impact equivalent to those set out in Article 51(1), point (a), the Commission shall take into account the following criteria:

(a) the number of parameters of the model;

(b) the quality or size of the data set, for example measured through tokens;

(c) the amount of computation used for training the model, measured in floating point operations or indicated by a combination of other variables such as estimated cost of training, estimated time required for the training, or estimated energy consumption for the training;

(d) the input and output modalities of the model, such as text to text (large language models), text to image, multi-modality, and the state of the art thresholds for determining high-impact capabilities for each modality, and the specific type of inputs and outputs (e.g. biological sequences);

(e) the benchmarks and evaluations of capabilities of the model, including considering the number of tasks without additional training, adaptability to learn new, distinct tasks, its level of autonomy and scalability, the tools it has access to;

(f) whether it has a high impact on the internal market due to its reach, which shall be presumed when it has been made available to at least 10 000 registered business users established in the Union;

(g) the number of registered end-users.

Recitals

Recital 111

It is appropriate to establish a methodology for the classification of general-purpose AI models as general-purpose AI model with systemic risks. Since systemic risks result from particularly high capabilities, a general-purpose AI model should be considered to present systemic risks if it has high-impact capabilities, evaluated on the basis of appropriate technical tools and methodologies, or significant impact on the internal market due to its reach. High-impact capabilities in general-purpose AI models means capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models. The full range of capabilities in a model could be better understood after its placing on the market or when deployers interact with the model. According to the state of the art at the time of entry into force of this Regulation, the cumulative amount of computation used for the training of the general-purpose AI model measured in floating point operations is one of the relevant approximations for model capabilities. The cumulative amount of computation used for training includes the computation used across the activities and methods that are intended to enhance the capabilities of the model prior to deployment, such as pre-training, synthetic data generation and fine-tuning. Therefore, an initial threshold of floating point operations should be set, which, if met by a general-purpose AI model, leads to a presumption that the model is a general-purpose AI model with systemic risks. This threshold should be adjusted over time to reflect technological and industrial changes, such as algorithmic improvements or increased hardware efficiency, and should be supplemented with benchmarks and indicators for model capability. To inform this, the AI Office should engage with the scientific community, industry, civil society and other experts. Thresholds, as well as tools and benchmarks for the assessment of high-impact capabilities, should be strong predictors of generality, its capabilities and associated systemic risk of general-purpose AI models, and could take into account the way the model will be placed on the market or the number of users it may affect. To complement this system, there should be a possibility for the Commission to take individual decisions designating a general-purpose AI model as a general-purpose AI model with systemic risk if it is found that such model has capabilities or an impact equivalent to those captured by the set threshold. That decision should be taken on the basis of an overall assessment of the criteria for the

designation of a general-purpose AI model with systemic risk set out in an annex to this Regulation, such as quality or size of the training data set, number of business and end users, its input and output modalities, its level of autonomy and scalability, or the tools it has access to. Upon a reasoned request of a provider whose model has been designated as a general-purpose AI model with systemic risk, the Commission should take the request into account and may decide to reassess whether the general-purpose AI model can still be considered to present systemic risks.

Recital 112

It is also necessary to clarify a procedure for the classification of a general-purpose AI model with systemic risks. A general-purpose AI model that meets the applicable threshold for high-impact capabilities should be presumed to be a general-purpose AI models with systemic risk. The provider should notify the AI Office at the latest two weeks after the requirements are met or it becomes known that a general-purpose AI model will meet the requirements that lead to the presumption. This is especially relevant in relation to the threshold of floating point operations because training of general-purpose AI models takes considerable planning which includes the upfront allocation of compute resources and, therefore, providers of general-purpose AI models are able to know if their model would meet the threshold before the training is completed. In the context of that notification, the provider should be able to demonstrate that, because of its specific characteristics, a GPAI model exceptionally does not present systemic risks, and that it thus should not be classified as a general-purpose AI model with systemic risks. That information is valuable for the AI Office to anticipate the placing on the market of general-purpose AI models with systemic risks and the providers can start to engage with the AI Office early on. That information is especially important with regard to general-purpose AI models that are planned to be released as open-source, given that, after the open-source model release, necessary measures to ensure compliance with the obligations under this Regulation may be more difficult to implement.

Recital 113

If the Commission becomes aware of the fact that a general-purpose AI model meets the requirements to classify as a general-purpose AI model with systemic risk, which previously had either not been known or of which the relevant provider has failed to notify the Commission, the Commission should be empowered to designate it so. A system of qualified alerts should ensure that the AI Office is made aware by the scientific panel of general-purpose AI models that should possibly be classified as general-purpose AI models with systemic risk, in addition to the monitoring activities of the AI Office.

Select bibliography

- Bernsteiner C and Schmitt T R, 'Art. 51 Einstufung von KI-Modellen mit allgemeinem Verwendungszweck als KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko' in Mario Martini and Christiane Wendehorst (eds), *KI-VO: Verordnung über Künstliche Intelligenz: Kommentar* (2nd edn, C H Beck 2026).
- Bomhard D and Siglmüller J, 'AI Act – das Trilogergebnis' (2024) Recht Digital 45.

- Bond T and Abbady S, 'Article 51: Classification of General-Purpose AI Models as General-Purpose AI Models with Systemic Risk' in Ceyhun Necati Pehlivan, Nikolaus Forgó and Peggy Valcke (eds), *The EU Artificial Intelligence (AI) Act: A Commentary* (Wolters Kluwer 2024).
- Carey S, 'Regulating Uncertainty: Governing General-Purpose AI Models and Systemic Risk' (2025) European Journal of Risk Regulation <https://doi.org/10.1017/err.2025.10040>.
- Erben A, Negele M, Heim L and Sevilla J, *Training Compute Thresholds - Key Considerations for the EU AI Act*, Fernández Llorca D, Gómez E (eds), (Publications Office of the European Union, JRC143255, 2025).
- Förster Chr, Straburzynski J, '§ 1 Grundlegende Begriffe und Konzepte der KI-VO' in Christian Förster (ed), *Die KI-Verordnung in der Praxis: Rechtliche Grundlagen und Pflichten bei der Anwendung von KI im Unternehmen* (C H Beck 2025).
- Haar T and Siglmüller J, 'Art. 51 Einstufung von KI-Modellen mit allgemeinem Verwendungszweck als KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko' in Jens Schefzig and Robert Kilian (eds), *Beck'scher Online-Kommentar KI-Recht* (4th edn, C H Beck 2025).
- Hilgendorf E and Härtlein J, 'Art. 51 Einstufung von KI-Modellen mit allgemeinem Verwendungszweck als KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko' in Eric Hilgendorf and Johannes Härtlein (eds.), *KI-VO: Verordnung über künstliche Intelligenz* (Nomos 2025).
- Hobbhahn M, Hovy D and Vanschoren J, *A Proposal to Identify High-Impact Capabilities in General-Purpose AI Models,* Fernández Llorca D, Gómez E (eds), (Publications Office of the European Union, JRC143258, 2025).
- Hofmann-Coombe J, '§ 7. KI-Modelle mit allgemeinem Verwendungszweck' in Eric Hilgendorf and David Roth-Isigkeit (eds), *Die neue Verordnung der EU zur Künstlichen Intelligenz* (2nd edn, C H Beck 2025).
- Martini M, '§ 3. Risikobasierter Ansatz' in Eric Hilgendorf and David Roth-Isigkeit (eds), *Die neue Verordnung der EU zur Künstlichen Intelligenz* (2nd edn, C H Beck 2025).
- Schneider A and Schneider L, 'Art. 51 Einstufung von KI-Modellen mit allgemeinem Verwendungszweck als KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko' in David Bomhard, Fritz-Ulli Pieper and Susanne Wende (eds), *Kommentar KI-VO: Verordnung über Künstliche Intelligenz* (Fachmedien Recht und Wirtschaft 2025).
- Schöbel Ph, Yang-Jacobi A M, 'Systemische Risiken im Zeitalter generativer KI' (2025) Recht Digital 627.
- Somala V, Ho A, Krier S, 'Three Challenges Facing Compute-based AI Policies' (2025) <https://epoch.ai/gradient-updates/three-issues-undermining-compute-based-ai-policies> accessed 7 January 2026.
- Vanschoren J, *The Role of AI Safety Benchmarks in Evaluating Systemic Risks in General-Purpose AI Models*, Fernández Llorca D, Gómez E (eds), (Publications Office of the European Union, JRC143259, 2025).

Commentary

1. General remarks

1.1. Introduction

1. Article 51 AI Act[1] sets out rules for the classification of general-purpose AI ("GPAI") models as GPAI models with systemic risk. These classification rules are the basis for the AI Act's two-tiered approach[2] to the regulation of GPAI models, with some obligations applying to providers of all

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 1689/1 ("AI Act").
[2] See Claudio Novelli and others, 'A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities' (2025) 16 European Journal of Risk Regulation 566, 572; David Bomhard and Jonas

GPAI models[3] and additional, more stringent, obligations applying for providers of GPAI models with systemic risk.[4] Article 51 only applies to GPAI models,[5] not to GPAI systems into which such models may be integrated.[6]

2. Article 52 and Annex XIII contain further provisions for the classification of GPAI models as presenting systemic risk.[7] Articles 51 and 52 together make up the 'Classification rules'[8] contained in Section 1 of Chapter V of the AI Act; their relationship is complex.[9] As Article 51(1) establishes substantive requirements for classification and Article 52 is entitled 'Procedure',[10] this could suggest a division whereby Article 51 contains substantive rules for classification and Article 52 contains

---

Siglmüller, 'AI Act – das Trilogergebnis' (2024) Recht Digital 45, para 29; Mario Martini, '§ 3. Risikobasierter Ansatz' in Eric Hilgendorf and David Roth-Isigkeit (eds), *Die neue Verordnung der EU zur Künstlichen Intelligenz* (2nd edn, C H Beck 2025) para 190; for a critique of this tiered approach, see Sandra Wachter, 'Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond' 26 Yale Journal of Law & Technology 3 (2024) 671, 697.

[3] See AI Act, art 53(1). However, article 53(2) provides a partial exception for providers of certain free and open-source models (see commentary on Article 53, paras 110–114 in this work).

[4] See AI Act, art 55(1); further, see Adrian Schneider and Leonie Schneider, 'Art. 51 Einstufung von KI-Modellen mit allgemeinem Verwendungszweck als KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko' in David Bomhard, Fritz-Ulli Pieper and Susanne Wende (eds), *Kommentar KI-VO: Verordnung über Künstliche Intelligenz* (Fachmedien Recht und Wirtschaft 2025) para 1; Gregory Smith and others, 'General-Purpose Artificial Intelligence (GPAI) Models and GPAI Models with Systemic Risk: Classification and Requirements for Providers' (RAND, 2024) <https://www.rand.org/pubs/research_reports/RRA3243-1.html>; Martini (n 2) para 190; European Commission, 'General-Purpose AI Models in the AI Act – Questions & Answers' <https://digital-strategy.ec.europa.eu/en/faqs/general-purpose-ai-models-ai-act-questions-answers> accessed 7 January 2026.

[5] Article 3(63) defines a GPAI model as 'an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market'. For an analysis of this definition see forthcoming commentary on Article 3(63) in this work.

[6] Clemens Bernsteiner and Thomas Rainer Schmitt, 'Art. 51 Einstufung von KI-Modellen mit allgemeinem Verwendungszweck als KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko' in Mario Martini and Christiane Wendehorst (eds), *KI-VO: Verordnung über Künstliche Intelligenz: Kommentar* (2nd edn, C H Beck 2026) para 5; Jason Hofmann-Coombe, '§ 7. KI-Modelle mit allgemeinem Verwendungszweck' in Eric Hilgendorf and David Roth-Isigkeit (eds), *Die neue Verordnung der EU zur Künstlichen Intelligenz* (2nd edn, C H Beck 2025) para 9. See article 3(66) which defines a GPAI system as 'an AI system which is based on a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems'. One may note, however, that article 51 continues to apply to the GPAI *model* even after its integration into an AI system (see AI Act, recital 97, ninth sentence).

[7] See Hofmann-Coombe (n 6) para 35.

[8] See AI Act, ch V, s 1, title.

[9] See Tobias Haar and Jonas Siglmüller, 'Art. 51 Einstufung von KI-Modellen mit allgemeinem Verwendungszweck als KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko' in Jens Schefzig and Robert Kilian (eds), *Beck'scher Online-Kommentar KI-Recht* (4th edn, C H Beck 2025) para 17 who conclude that the relationship between Article 51 and 52 and the significance of Article 52 remains largely unclear; see also Hofmann-Coombe (n 6) paras 35–36; Adrian Schneider and Leonie Schneider, 'Art. 52 Verfahren' in David Bomhard, Fritz-Ulli Pieper and Susanne Wende (eds), *Kommentar KI-VO: Verordnung über Künstliche Intelligenz* (Fachmedien Recht und Wirtschaft 2025) para 13.

[10] For the role of the wording of an article's title in the interpretation of operative provisions, see Case C-311/18 *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems* [2020] ECLI:EU:C:2020:559 ("Schrems II") para 92; see also Case C-291/13 *Sotiris Papasavvas v O Fileleftheros Dimosia Etaireia Ltd and Others* [2014] ECLI:EU:C:2014:2209 ("Papasavvas") para 39 with regard to a section title.

procedural rules for classification.[11] The recitals imply a similar delineation between these provisions, suggesting that Article 51 was intended to 'establish a methodology', whereas Article 52 was intended to 'clarify a procedure' for the classification of GPAI models with systemic risk.[12] However, these distinctions arguably oversimplify the matter as the distinction is not as clear-cut.[13] Upon closer inspection, Article 51 also contains procedural provisions. For example, Article 51(1)(b) establishes the requirement of a Commission decision for the classification of models with capabilities or an impact equivalent to high-impact capabilities.[14] Conversely, Article 52 also contains provisions that concern substantive criteria for classification.[15] For example, Article 52(2) and (3) establish the substantive requirements under which the Commission can reject the arguments submitted by a provider to contest classification, doing so without reference to Article 51.[16] Overall, the close connection between the two provisions, which is reflected not least in the explicit references in Article 52(1) and (2) to Article 51(1)(a), makes a systematic interpretation of both provisions in context inevitable.[17]

3. Regarding the practical scope of these classification rules, the AI Act does not impose a strict limit to the number of GPAI models that may be classified as presenting systemic risk at a time. However, the AI Act's definition of systemic risk as being specific to capabilities that match or exceed the capabilities of the most advanced GPAI models,[18] the initial setting of Article 51(2)'s training compute threshold at 10^{25} floating-point operations ("FLOPs"),[19] and the Commission's duty to update this threshold 'in light of evolving technological developments'[20] all suggest that the legislature intended only a limited number of providers of GPAI models with systemic risk to be classified at

[11] See Martini (n 2) para 196; Eric Hilgendorf and Johannes Härtlein, 'Art. 52 Verfahren' in Eric Hilgendorf and Johannes Härtlein (eds.), *KI-VO: Verordnung über künstliche Intelligenz* (Nomos 2025) para 1; Lukas Feiler, Nikolaus Forgó and Michaela Nebel, 'Article 52' in *The EU AI Act: A Commentary* (Globe Law and Business 2025) para 1. The wording of an article's title plays a role in the interpretation of operative provisions (*Schrems II* (n 10) para 92; see also *Papasavvas* (n 10) para 39 with regard to a section's title), as long as the title is not provided for ease of reference only (Case C-97/15 *Sprengen/Pakweg Douane BV v Staatssecretaris van Financiën* [2016] ECLI:EU:C:2016:556 para 31 and the case law cited therein).

[12] AI Act, recital 111, first sentence, and recital 112, first sentence. As the delineation between methodology and procedure is nebulous, this offers little interpretive guidance. Moreover, while the recitals 'constitute important elements for the purposes of interpretation', they lack binding legal force (see, for example, Case C-418/18 *Patrick Grégor Puppinck and Others v European Commission* [2019] ECLI:EU:C:2019:1113 ("Puppinck") paras 75–76).

[13] See Haar and Siglmüller, 'Art. 51' (n 9) paras 17–22; Hilgendorf and Härtlein, 'Art. 52' (n 11) para 1.

[14] See AI Act, art 51(1)(b) ('based on a decision of the Commission, ex officio or following a qualified alert from the scientific panel') and Section 2.1.2.2. The omission of the hyphen in Article 51(1)(a) and (2)'s spelling of 'high impact capabilities' appears to be an unintended drafting inconsistency without substantive relevance (see Section 2.1.1.). Accordingly, this chapter adopts a spelling of the term as 'high-impact capabilities' in accordance with the term's spelling elsewhere in the AI Act.

[15] See Hilgendorf and Härtlein, 'Art. 52' (n 11) para 1.

[16] See commentary on Article 52, Section 2.2.2. in this work.

[17] Clemens Bernsteiner and Thomas Rainer Schmitt, 'Art. 52 Verfahren' in Mario Martini and Christiane Wendehorst (eds), *KI-VO: Verordnung über Künstliche Intelligenz: Kommentar* (2nd edn, C H Beck 2026) paras 4–6; Hofmann-Coombe (n 6) para 35; see also Haar and Siglmüller, 'Art. 51' (n 9) paras 18–22.

[18] AI Act, arts 3(65) and (64). For a discussion of the meaning of 'specific to' and 'most advanced' see forthcoming commentary on Article 3(65) in this work and the forthcoming commentary on Article 3(64) in this work respectively.

[19] It has been estimated that at the start of 2024 only four models existed that surpass this threshold, and at the start of 2025 seventeen models (see Ben Cottier and David Owen, 'How Many AI Models Will Exceed Compute Thresholds?' (2025) <https://epoch.ai/blog/model-counts-compute-thresholds#results> accessed 7 January 2026).

[20] AI Act, art 51(3) and recital 179, seventh sentence; see Section 2.3.2.

a time.[21] Indeed, the Safety and Security Chapter[22] of the Code of Practice was drafted assuming no more than fifteen providers would be subject to the obligations for GPAI models with systemic risk at a time.[23]

4. The rules for the classification of GPAI models as presenting systemic risk were adopted in light of Recitals 111–113.[24] These rules are further referred to in Recitals 163,[25] 173[26] and 179[27].

1.2. Structure & overview

5. This chapter contains a paragraph-by-paragraph analysis of Article 51. The substantive analysis begins in Section 2.1. with Article 51(1), which establishes two alternative conditions for systemic risk classification of GPAI models. The first condition requires a GPAI model to have high-impact capabilities evaluated through appropriate technical tools and methodologies.[28] The second condition requires the GPAI model to display capabilities or an impact equivalent to high-impact capabilities, determined through a Commission decision based on the criteria in Annex XIII. In particular, this section examines whether classification under Article 51(1)(a) occurs automatically or requires Commission designation[29] and considers what kind of capabilities or impact of a GPAI model could lead to its classification under Article 51(1)(b).[30] It also addresses questions closely

[21] However, see Philipp Hacker and Matthias Holweg, 'The Regulation of Fine-Tuning: Federated Compliance for Modified General-Purpose AI Models' (2026) 60 Computer Law & Security Review 106234 5–6 who argue against GPAI models *automatically* falling outside the 'systemic risk category' upon the release of new models, defending a 'static approach' to interpreting Article 3(64)'s definition of high-impact capabilities which 'would treat "most advanced" models as those identified as most advanced at the time of the AI Act's enactment (August 2024), or those surpassing a defined capability threshold that remains relatively stable over time' over a 'dynamic interpretation' which 'continually categorizes only the top few models' (without discussing the interplay of Article 3(64) with systemic risk classification under Section 1 of Chapter V of the AI Act).

[22] See European Commission, 'Code of Practice for General-Purpose AI Models – Safety and Security Chapter' (2025) <https://ec.europa.eu/newsroom/dae/redirection/document/118119> accessed 7 January 2026.

[23] 'Statement from the Chairs and Vice Chairs Responsible for the Drafting of the Safety and Security Chapter of the Code of Practice', <https://code-of-practice.ai/?section=safety-security#chair-statement> accessed 7 January 2026; see also European Commission, 'General-Purpose AI Models in the AI Act – Questions & Answers' (n 4), interpreting the reference to the most advanced model in Article 3(65)'s systemic risk definition as referring to the state-of-the-art.

[24] Haar and Siglmüller, 'Art. 51' (n 9) para 5; Tobias Haar and Jonas Siglmüller, 'Art. 52 Verfahren' in Jens Schefzig and Robert Kilian (eds), *Beck'scher Online-Kommentar KI-Recht* (4th edn, C H Beck 2025) para 2.

[25] See AI Act, recital 163: 'With a view to complementing the governance systems for general-purpose AI models, the scientific panel should support the monitoring activities of the AI Office and may, in certain cases, provide qualified alerts to the AI Office which trigger follow-ups, such as investigations. [...] Furthermore, this should be the case where the scientific panel has reason to suspect that a general-purpose AI model meets the criteria that would lead to a classification as general-purpose AI model with systemic risk.'

[26] See AI Act, recital 173, first sentence: 'In order to ensure that the regulatory framework can be adapted where necessary, the power to adopt acts in accordance with Article 290 TFEU should be delegated to the Commission to amend [...] the threshold, benchmarks and indicators, including by supplementing those benchmarks and indicators, in the rules for the classification of general-purpose AI models with systemic risk, the criteria for the designation of general-purpose AI models with systemic risk [...].'

[27] See AI Act, recital 179, seventh sentence: 'The AI Office should ensure that classification rules and procedures are up to date in light of technological developments.'

[28] See Section 2.1.1.

[29] See Section 2.1.1.1.

[30] See Section 2.1.2.1.3. and Section 2.1.2.1.4. respectively.

related to Article 51(1)'s classification conditions, such as the existence of further classification pathways outside of Article 51(1)[31] and the effects of classification.[32]

6. Section 2.2. addresses Article 51(2)'s presumption of high-impact capabilities based on the 10^{25} FLOPs training compute threshold. This section explores in particular the notion of 'cumulative amount of computation used for [the model's] training' – determining what computational activities qualify for inclusion and methods for estimating the amount of training compute.[33]

7. Section 2.3. examines Article 51(3)'s delegation of power to the Commission to amend thresholds and supplement benchmarks and indicators. It pays particular attention to interpretative questions regarding the scope of this delegation of power, including whether it extends to substantive classification criteria under Article 51(1)[34] and whether the Commission is obliged to exercise these powers when necessary to reflect technological developments.[35]

8. The chapter concludes in Section 2.4. with a brief discussion of the relevance of Annex XIII for classification.[36]

9. Moreover, it is key to emphasise that Article 51 is interconnected with numerous other provisions of the AI Act, including the definitions of a general-purpose AI model, of high-impact capabilities and of systemic risk under Article 3,[37] and the Article 52 rules on 'Procedure',[38] and cannot be interpreted in isolation. Where appropriate, the analysis therefore provides cross-references to other chapters of this Commentary addressing these provisions and the respective interpretive questions they pose.

[31] See Section 2.1.3.
[32] See Section 2.1.4.
[33] See Section 2.2.1.
[34] See Section 2.3.1.
[35] See Section 2.3.2.
[36] See Section 2.4.
[37] See AI Act, art 3(63)–(65). These definitions are analysed in the forthcoming commentary on Article 3(63), the forthcoming commentary on Article 3(64) and the forthcoming commentary on Article 3(65) in this work respectively.
[38] See commentary on Article 52 in this work.

2. Substance

2.1. Article 51(1): Classification of GPAI models as presenting systemic risk

10. Article 51(1) introduces two *alternative* conditions under which a GPAI model shall be classified as presenting systemic risk.[39] The first condition, under point (a),[40] requires the model to exhibit high-impact capabilities. The second condition, under point (b),[41] concerns the presence of capabilities or an impact equivalent to those high-impact capabilities, as determined by the Commission on the basis of the criteria set out in Annex XIII.

11. In addition to Article 51(1), Article 52 sets out the Commission's power to designate GPAI models as presenting systemic risk,[42] thereby raising the question of the relationship of Article 52's designation provisions with the classification conditions under Article 51(1). While this relationship is analysed in-depth elsewhere,[43] it is apparent that Article 51(1) was intended to serve as the foundational provision for the classification of a GPAI model as presenting systemic risk, as evidenced by the wording of the provision and its positioning at the beginning of the classification rules within Section 1 of Chapter V of the AI Act.[44] The central importance of Article 51(1) is reflected in the Commission's Guidelines on the Scope of the Obligations for General-Purpose AI Models ("Commission Guidelines"), which state that '[f]rom the moment when a general-purpose AI model meets either of the two conditions [under Article 51(1)], the model is classified as a general-purpose AI model with systemic risk and its provider must comply with the relevant obligations'.[45]

2.1.1. Article 51(1)(a)

12. Article 51(1)(a) sets out the first alternative condition for the classification of a GPAI model as presenting systemic risk, requiring that the model 'has high impact capabilities evaluated on the

[39] European Commission, 'Communication from the Commission - Commission Guidelines on the Scope of the Obligations for Providers of General-Purpose AI Models Established by Regulation (EU) 2024/1689 (AI Act)' C(2025) 7719 final ("Commission Guidelines") paras 25–26; European Commission, 'General-Purpose AI Models in the AI Act – Questions & Answers' (n 4); Moritz Hecht, 'Regulierung von GPAI-Modellen durch die KI-Verordnung' (2025) Künstliche Intelligenz und Recht 30, 33; Smith and others (n 4); Christian Förster and Julia Straburzynski, '§ 1 Grundlegende Begriffe und Konzepte der KI-VO' in Christian Förster (ed), *Die KI-Verordnung in der Praxis: Rechtliche Grundlagen und Pflichten bei der Anwendung von KI im Unternehmen* (C H Beck 2025) para 66. The view that both conditions under Article 51 come with the same substantive requirements for classification (see Bernsteiner and Schmitt, 'Art. 51' (n 6) para 25); see also Martini (n 2) para 192; Andreas Engel, 'Generative KI, Foundation Models und KI-Modelle mit allgemeinem Verwendungszweck in der KI-VO: Passende Mosaiksteine?' (2024) Künstliche Intelligenz und Recht 21, 23) fails to recognise the legislative decision to establish two alternative classification conditions (see AI Act, art 51(1): 'if it meets any of the following conditions') and is incompatible with the wording of the provision as discussed below (see Section 2.1.2.1.1.).

[40] See Section 2.1.1.

[41] See Section 2.1.2.

[42] AI Act, art 52(1), first sentence, and art 52(4), first subparagraph.

[43] See commentary on Article 52, Section 2.1.3.1. in this work; see also the commentary on Article 52, Section 2.1.3.3. and Section 2.3.1.1. in this work.

[44] See AI Act, art 51(1): 'A general-purpose AI model shall be classified as a general-purpose AI model with systemic risk *if it meets any of the following conditions*: [...]' (emphasis added).

[45] Commission Guidelines (n 39) para 28; for the effects of classification, see also Section 2.1.4.

basis of appropriate technical tools and methodologies, including indicators and benchmarks'.[46] Article 3(64) defines high-impact capabilities as 'capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models'.[47]

13. The classification of GPAI models with high-impact capabilities as presenting systemic risk under Article 51(1)(a) is best understood within the overall context of the classification rules under Articles 51 and 52.[48] Under Article 51(2), a model is presumed to have high-impact capabilities, once the cumulative amount of computation used for its training measured in FLOPs is greater than 10^{25}. Where a GPAI model has high-impact capabilities or it becomes known that it will have high-impact capabilities, its provider has to notify the Commission pursuant to Article 52(1)'s first sentence.[49] Further, Article 52(2) allows the provider to contest the classification of the model when submitting the notification.[50] Moreover, Article 52(1)'s third sentence allows the Commission to designate a GPAI model of which it has not been notified as 'presenting systemic risks' as a GPAI model with systemic risk.[51] This designation provision is arguably linked to Article 51(1)(a) as well.[52] While Annex XIII primarily informs the determination of whether a GPAI model may be classified under Article 51(1)(b)'s condition,[53] certain criteria in the annex may also inform the assessment of high-impact capabilities under Article 51(1)(a),[54] particularly those included in points (d) and (e) of Annex XIII.[55]

14. A thorough understanding of Article 51(1)(a) requires examining the relationship between high-impact capabilities and systemic risk. Article 3(65) defines systemic risk as 'a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain'.[56] In this context, Article 51(1)(a) can be read as creating a kind of presumption of systemic risk presence in a GPAI model based on its high-impact capabilities.[57] This finds support in Recital 112's first sentence, which states that '[a] general-purpose AI model that meets the applicable threshold for high-impact capabilities should be presumed to be a general-purpose AI model with systemic risk.' The procedure to contest classification under Article 52's

[46] See Haar and Siglmüller, 'Art. 51' (n 9) para 32; Schneider and Schneider, 'Art. 51' (n 4) 11; see also AI Act, recital 111, second sentence.

[47] For interpretive issues posed by this definition, see Section 2.1.1.2.; for an analysis of this definition, see forthcoming commentary on Article 3(64) in this work.

[48] See Hofmann-Coombe (n 6) paras 44, 47; for the relationship between articles 51 and 52, see also Section 1.1.

[49] See commentary on Article 52, Section 2.1.1.2. in this work.

[50] See commentary on Article 52, Section 2.2. in this work.

[51] See commentary on Article 52, Section 2.1.3. in this work.

[52] See commentary on Article 52, Section 2.1.3.1. in this work.

[53] See Section 2.1.2.1.2.

[54] See Section 2.4.2.; for an overview over Annex XIII's criteria, see Section 2.4.1.

[55] See AI Act, annex XIII, point (d) ('the input and output modalities of the model, such as text to text (large language models), text to image, multi-modality, and the state of the art thresholds for determining high-impact capabilities for each modality, and the specific type of inputs and outputs (e.g. biological sequences)') and point (e) ('the benchmarks and evaluations of capabilities of the model, including considering the number of tasks without additional training, adaptability to learn new, distinct tasks, its level of autonomy and scalability, the tools it has access to'); see also the forthcoming commentary on Article 3(64) in this work.

[56] For a discussion of this definition, see forthcoming commentary on Article 3(65) in this work.

[57] See Haar and Siglmüller, 'Art. 51' (n 9) para 34 who argue and criticize that Article 51(1)(a) infers the existence of systemic risk from a model's high-impact capabilities (without expressly characterising Article 51(1)(a) as a presumption).

second and third paragraph would then operate as a mechanism to rebut that presumption.[58] However, unlike Article 51(2),[59] Article 51(1)(a) does not expressly characterise the link between high-impact capabilities and systemic risk as 'presumptive' and Article 52(2) and (3) do not expressly characterise the procedure to contest classification as a 'rebuttal' of a presumption.[60]

15. The contestation procedure under Article 52(2) and (3) explains why a GPAI model's classification under Article 51(1)(a) generally does not require an independent determination of whether it comes with risks that meet Article 3(65)'s definition of systemic risk.[61] Rather, where a model has high-impact capabilities but exceptionally does not present systemic risks, the provider can contest the model's classification pursuant to Article 52(2).[62] A similar rationale applies to the assessment of whether a model has a significant impact on the internal market:[63] although this criterion is expressly mentioned in Article 3(65)'s definition of systemic risk[64] as well as in Annex XIII,[65] Article 51(1)(a)'s wording unambiguously precludes an independent requirement that the model should have a significant impact on the internal market under this provision.[66] Rather, this assessment may be relevant to the procedure for contesting classification under Article 52(2), as well as designation under Article 51(1)(b).[67] Similar considerations also apply for other criteria without direct relevance

[58] For this procedure see commentary on Article 52, Section 2.2.2.1. in this work.

[59] See AI Act, art 51(2): 'A general-purpose AI model *shall be presumed* to have high-impact capabilities pursuant to paragraph 1, point (a), when [...]'. (emphasis added)

[60] The Commission Guidelines notably do not use the word 'presumption' with regards to Article 51(1)(a)'s classification condition either (see Commission Guidelines (n 39) para 39).

[61] See Haar and Siglmüller, 'Art. 51' (n 9) para 32; Martini (n 2) para 196; Philipp Schöbel and Anna Maria Yang-Jacobi, 'Systemische Risiken im Zeitalter generativer KI' (2025) Recht Digital 627, 632; opposing view: Philipp Hacker, Atoosa Kasirzadeh and Lilian Edwards, 'AI, Digital Platforms, and the New Systemic Risk' (2025) <https://arxiv.org/abs/2509.17878> accessed 7 January 2026, 16.

[62] Martini (n 2) para 196. As the provider may also refrain from contesting classification in such a case, this suggests that the AI Act allows for the systemic risk classification of a GPAI model which does not come with systemic risks as defined under Article 3(65). This potential 'overinclusiveness' of the classification rules under Section 1 of Chapter V is not necessarily problematic in light of the obligations that follow from classification. In particular, Article 55(1)(a) and (b) oblige providers to perform model evaluations with a view to '*identifying* [...] systemic risks' (emphasis added) and to 'assess and mitigate *possible* systemic risks [...] that *may* stem from the [GPAI model]' (emphasis added). Thus, these obligations are phrased in a way that does not necessarily require the actual presence of systemic risk as defined under Article 3(65). A rather overinclusive approach to systemic risk classification aligns with the precautionary principle as a general principle of EU law that 'implies that where there is uncertainty as to the existence or extent of risks to human health, the institutions may take precautionary measures without having to wait until the reality and seriousness of those risks becomes fully apparent' (Cases T-74, 76, 83-85, 132, 137 and 141/00 *Artegodan GmbH and Others v Commission of the European Communities* [2002] ECR II-4945 paras 184–185).

[63] See Haar and Siglmüller, 'Art. 51' (n 9) para 33; opposing view: Bernsteiner and Schmitt, 'Art. 51' (n 6) para 29.

[64] See AI Act, art 3(65): '"systemic risk" means a risk that is specific to the high-impact capabilities of general-purpose AI models, *having a significant impact on the Union market* due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain;' (emphasis added). For an analysis of whether this definition relates the 'significant impact on the Union market' to GPAI models or their risk, see forthcoming commentary on Article 3(65) in this work.

[65] See AI Act, annex XIII: 'For the purpose of determining that a general-purpose AI model has capabilities or an impact equivalent to those set out in Article 51(1), point (a), the Commission shall take into account the following criteria: [...] (f) *whether it has a high impact on the internal market* due to its reach, which shall be presumed when it has been made available to at least 10 000 registered business users established in the Union;' (emphasis added).

[66] Haar and Siglmüller, 'Art. 51' (n 9) para 33.

[67] See Section 2.1.2.1.4.

for the model's capabilities, such as the number of registered end users mentioned under point (g) of Annex XIII.[68]

2.1.1.1. Commission designation in the context of Article 51(1)(a)

16. Scholars are divided on whether the classification of GPAI models with high-impact capabilities under Article 51(1)(a) requires a Commission decision[69] or whether GPAI models meeting that condition are automatically classified by operation of law.[70] According to its Guidelines, the Commission does not view designation as a necessary requirement for classification under Article 51(1)(a), setting out that a provider must comply with the obligations for providers of GPAI models with systemic risk from 'the moment when the model meets the condition laid down in Article 51(1), point (a), AI Act'.[71]

17. An interpretation of Article 51(1)(a) as providing for the automatic classification, by operation of law, of GPAI models with high-impact capabilities as presenting systemic risk is compelling, as the legislative text – despite some ambiguity in the wording of Article 51(1) – more strongly supports it. Some scholars have argued – based on the German language version of Article 51(1) – that the wording 'shall be classified' instead of 'is classified' or 'is considered' implies a further procedural requirement for classification beyond Article 51(1)(a)'s substantive requirement of high-impact capabilities.[72] However, this interpretation is largely undermined by Article 51(1)(b), which expressly requires that determination of equivalent capabilities or impact must be 'based on a decision of the Commission'. The express provision for a Commission decision in Article 51(1)(b)'s classification condition indicates that no such requirement exists for Article 51(1)(a).[73] This interpretation is reinforced by Recital 111's second sentence, which states that 'a general-purpose AI model should be considered to present systemic risks if it has high-impact capabilities' without mentioning the requirement of a Commission decision in this context.

18. A comparison with the Digital Markets Act ("DMA") and the Digital Services Act ("DSA") reinforces this reading.[74] Article 51(1)'s use of 'shall be classified' notably deviates both from the use of 'shall be designated' in the DMA's provisions for gatekeeper designation and the use of 'shall [...] adopt a decision designating as [...]' in the DSA's provisions for designation of very large online

[68] Opposing view: Theodoros Karathanis, 'Fitting "Systemic Risks" into a Taxonomy in the GPAI Code of Practice: Will the Resulting Ambiguity be Exploited by GPAI Model Providers?' (2025) 28 Journal of Internet Law 6, 11.

[69] In favour of the requirement of a Commission decision for classification under article 51(1)(a): Haar and Siglmüller, 'Art. 51' (n 9) paras 26–31; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 12; Hofmann-Coombe (n 6) para 35; Hilgendorf and Härtlein, 'Art. 52' (n 11) para 3; Martini (n 2) para 197; Schöbel and Yang-Jacobi (n 61) 632; Claudio Novelli and others, 'Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity', 55 Computer Law & Security Review (2024) 106066, 2–3.

[70] In favour of automatic classification under Article 51(1)(a): Toby Bond and Shima Abbady, 'Article 52 Procedure' in Ceyhun Necati Pehlivan, Nikolaus Forgó and Peggy Valcke (eds), *The EU Artificial Intelligence (AI) Act: A Commentary* (Wolters Kluwer 2024) 839–840, sec 1; see also Förster and Straburzynski (n 39) para 66.

[71] Commission Guidelines (n 39) para 26; see also Commission Guidelines (n 39) paras 41, 44.

[72] Haar and Siglmüller, 'Art. 51' (n 9) para 27. In the German language version of Article 51(1) 'shall be classified' is translated as 'wird [...] eingestuft'.

[73] Further, an interpretation of 'shall be classified' as requiring a Commission decision would arguably imply that classification under Article 51(1)(b) requires two Commission decisions – as the requirement of a Commission decision is already mentioned in Article 51(1)(b) itself – which does not appear close at hand (however, see Hofmann-Coombe (n 6) paras 48, 51 who argues for a two-stage procedure in the context of Article 51(1)(b)).

[74] See Bond and Abbady, 'Art. 52' (n 70) 839–840 s 1 with regard to the DSA.

platforms ("VLOPs") and very large online search engines ("VLOSEs").[75] This terminological difference between Article 51(1) on one hand and Articles 3(1) DMA and 33(4) DSA on the other hand can plausibly be explained by the AI Act's need to accommodate both for classification by operation of law under Article 51(1)(a) and classification by Commission decision under Article 51(1)(b).

19. Beyond arguments grounded in Article 51(1)'s wording, the apparent lack of a legal basis for a designation decision where the provider has notified the Commission of a model's high-impact capabilities further supports automatic classification under Article 51(1)(a). While Article 52 does contain two designation provisions, neither of those allow the Commission to designate a GPAI model of which it has been notified solely on the basis of its high-impact capabilities. In fact, Article 52(1)'s third sentence permits designation only of GPAI models 'of which [the Commission] *has not been notified*',[76] while Article 52(4)'s first subparagraph addresses the designation of a GPAI based on the criteria set out in Annex XIII – and not on the model's high-impact capabilities.[77] This, too, strongly suggests that Article 51(1)(a) does not require a Commission decision for classification of a GPAI model with high-impact capabilities as a GPAI model with systemic risk.

20. The counterarguments against automatic classification under Article 51(1)(a) are not entirely convincing. Some authors contend that the vagueness of Article 51(1)'s substantive requirements and the complexity of their assessment causes legal uncertainty for providers that would not permit automatic classification.[78] This concern appears reasonable with respect to classification under Article 51(1)(b) which generally requires 'an overall assessment' of the criteria set out in Annex XIII[79] – and can indeed prove challenging to perform for providers. However, the argument is far less strong for classification under Article 51(1)(a) based on a model's high-impact capabilities. For this classification condition, Article 51(2) provides a presumption of high-impact capabilities based on the cumulative amount of computation used for training. Given the existing methods for estimating training compute[80] and the Commission Guidelines' recognition that some leeway is

[75] See Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) [2022] OJ L 265/1 ("DMA"), art 3(1), and Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L 277/1 ("DSA"), art 33(4). Haar and Siglmüller, 'Art. 51' (n 9) para 28 argue that the legislative process leading up to the AI Act reflected agreement to design the rules for classification of GPAI models as presenting systemic risk parallel to the designation procedure established under Articles 33 DSA. Since that procedure requires a Commission decision under Article 33(4) DSA, they argue, such a decision must likewise be necessary for Article 51(1)(a). However, this argument is countered by the fact that the actual provisions for classification of GPAI models as presenting systemic risk under Articles 51 and 52 deviate textually from the rules for designation of VLOPs and VLOSEs under Article 33 DSA (see Bond and Abbady, 'Art. 52' (n 70) 839–840 s 1). In particular, Articles 51 and 52 use the term 'classification', a term not used in the DSA. For a general discussion of the different legal instruments from which the AI Act appears to have drawn inspiration and the implications for analogical interpretation, see forthcoming chapter on Common Legal Arguments in this work.

[76] See commentary on Article 52, Section 2.1.3.2. in this work.

[77] See commentary on Article 52, Section 2.3.1.1. in this work.

[78] Haar and Siglmüller, 'Art. 51' (n 9) para 27; see also Bernsteiner and Schmitt, 'Art. 52' (n 17) para 12 (arguing that a provider has a legitimate interest in obtaining clarity about the obligations to which it is subject).

[79] See AI Act, recital 111, twelfth sentence; see also AI Act, art 51(1)(b), annex XIII.

[80] See Commission Guidelines (n 39) paras 120–130; Alexander Erben and others, 'Training Compute Thresholds - Key Considerations for the EU AI Act' (Publications Office of the European Union, JRC143255, 2025) <https://publications.jrc.ec.europa.eu/repository/handle/JRC143255> 30–32; Jaime Sevilla and others, 'Estimating Training Compute of Deep Learning Models' (2022) <https://epoch.ai/blog/estimating-training-compute> accessed 7 January 2026.

appropriate in making estimates, to 'account for the difficulties providers may face',[81] a provider can determine without excessive difficulty whether its model meets the training compute threshold.[82] Moreover, other provisions, such as the notification obligation under Article 52(1)'s first and second sentence, demonstrate that the AI Act expects providers to self-assess whether their models have high-impact capabilities.[83] Article 51(1)(a) and (3) further envisage the Commission's adoption of indicators and benchmarks to help evaluate the model's high-impact capabilities via delegated act,[84] which may further facilitate such assessments in the future.

21. A second argument against automatic classification under Article 51(1)(a) posits that the possibility to contest classification under Article 52(2) can be plausibly explained only if classification under Article 51(1)(a) requires a Commission designation.[85] This argument relies on the fact that the AI Act does not expressly provide an exemption from classification if the Commission accepts a provider's arguments submitted pursuant to Article 52(2) to contest classification.[86] While Article 52(3) expressly addresses only the rejection, and not the acceptance, of a provider's arguments and its consequences, this omission, however, does not support the conclusion that classification under Article 51(1)(a) requires designation. The necessity of an acceptance decision is independent of any requirement for a designation decision for classification under Article 51(1)(a).[87] Moreover, Article 52(3) does not empower the Commission to designate a GPAI model where the provider's contestation of classification fails to meet the requisite standard. Rather, in such cases the Commission 'reject[s]' the provider's arguments and the GPAI model 'shall be considered' to be a GPAI model with systemic risk.[88] Furthermore, there appears to be no basis to infer a requirement for Commission designation under Article 51(1)(a)'s classification condition from the absence of provisions governing the acceptance of a provider's contestation pursuant to Article 52(2), particularly given that – as demonstrated above – such a decision lacks an express legal basis.

22. Even though, proceeding from the above, classification under Article 51(1)(a) does not require a Commission designation decision, the Commission can nevertheless designate a GPAI model with high-impact capabilities as a GPAI model with systemic risk under Article 52(1)'s third sentence when it has not been notified of it.[89] Although this provision refers to a 'general-purpose AI model *presenting systemic risks*' (emphasis added), there are convincing arguments to read designation under Article 52(1)'s third sentence in the context of the rest of Article 52(1), which also relates to GPAI models with high-impact capabilities – as laid out in further detail elsewhere.[90]

---

[81] See Commission Guidelines (n 39) para 120 according to which '[p]roviders may choose any method to estimate the relevant amount of training compute, so long as the estimated amount is, in the providers' best judgment, accurate within an overall error margin of 30% of the reported estimate'; see also Section 2.2.1.3.1.

[82] See also AI Act, recital 112, fourth sentence ('[T]training of general-purpose AI models takes considerable planning which includes the upfront allocation of compute resources and, therefore, providers of general-purpose AI models are able to know if their model would meet the threshold before training is completed.').

[83] In particular, article 52(1)'s second sentence requires providers to include information to demonstrate that the requirements for notification are met in the notification (see commentary on Article 52, Section 2.1.2. in this work) which presupposes the provider's prior assessment to that effect.

[84] See Section 2.3.1.3.

[85] Haar and Siglmüller, 'Art. 51' (n 9) para 28; see also Hilgendorf and Härtlein, 'Art. 52' (n 11) para 4 (arguing on the basis of article 52(3) that, in case of a reasoned challenge to classification under article 52(2), the Commission does not grant an exemption from classification but merely refrains from classifying the model).

[86] Haar and Siglmüller, 'Art. 51' (n 9) para 28; see also Hilgendorf and Härtlein, 'Art. 52' (n 11) para 4.

[87] See commentary on Article 52, Section 2.2.3.1. in this work.

[88] AI Act, art 52(3); see commentary on Article 52, Section 2.2.3.2. in this work.

[89] See commentary on Article 52, Section 2.1.3.3. in this work.

[90] See commentary on Article 52, Section 2.1.3. in this work.

2.1.1.2. High-impact capabilities

23. While discussed in-depth elsewhere,[91] Article 3(64) is a key provision to thoroughly understand the concept of high-impact capabilities. It defines them as 'capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models'. The omission of the hyphen in Article 51(1)(a) and (2)'s spelling of 'high impact capabilities' appears to be an unintended drafting inconsistency without substantive relevance that does not prevent the definition of high-impact capabilities under Article 3(64) from applying in the context of Article 51(1)(a).[92]

24. Article 3(64)'s definition raises several interconnected interpretive questions that are touched upon throughout this chapter. For example, what constitutes the most advanced GPAI models – is this determined domain-specifically (such as most advanced in coding or video generation) or by overall advancement,[93] and if the latter, how would overall advancement be assessed? Once the most advanced GPAI models are identified, which capabilities should be compared – must a model match or exceed all capabilities or only relevant ones, and which capabilities would be relevant?[94] Does Article 3(64)'s reference to 'recorded' capabilities further limit the capabilities under consideration?[95]

2.1.1.3. Appropriate technical tools and methodologies, including indicators and benchmarks

25. Whether a GPAI model has high-impact capabilities must be evaluated on the basis of 'appropriate technical tools and methodologies, including indicators and benchmarks'.[96] This enumeration must be interpreted in light of the AI Act's reference to a multitude of such assessment instruments in the context of systemic risk classification, such as '(technical) tools',[97] 'methodologies',[98] 'indicators',[99]

---

[91] See forthcoming commentary on Article 3(64) in this work.

[92] While article 3(64) and (65) spells the term as 'high-impact capabilities' with a hyphen, article 51(1)(a) and (2) omit the hyphen, rendering it as 'high impact capabilities'. This inconsistency extends to recitals 111 and 112, which relate to article 51 but retain the hyphenated spelling. The drafting inconsistency may explain the existence of two different translations of high(-)impact capabilities in the German language version of the AI Act: 'Fähigkeiten mit hoher Wirkkraft' under article 3(64) and (65) and 'Fähigkeiten mit hohem Wirkungsgrad' under article 51(1)(a) and (2). Based on available sources, no significance is attributed to this distinction based either on the German language version or on other language versions (see Commission Guidelines (n 39) para 25; Hofmann-Coombe (n 6) para 38; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 24).

[93] See Section 2.1.2.1.3.

[94] See Section 2.1.2.1.3. and Section 2.1.2.1.3.2.

[95] See Section 2.1.2.1.3.

[96] AI Act, art 51(1)(a) and recital 111, second sentence; for a recent proposal of how to assess whether a GPAI model has high-impact capabilities based on principal component analysis ("PCA") from a model's results on a selection of benchmarks, see Marius Hobbhahn and others, 'A Proposal to Identify High-Impact Capabilities in General-Purpose AI Models' (Publications Office of the European Union, JRC143258, 2025) <https://op.europa.eu/en/publication-detail/-/publication/65908a6e-a585-11f0-a7c5-01aa75ed71a1/language-en>; for an overview over capability thresholds as an approximation for risk associated with frontier AI, see Leonie Koessler, Jonas Schuett and Markus Anderljung, 'Risk Thresholds for Frontier AI' (2024) <https://arxiv.org/abs/2406.14713> accessed 7 January 2026, 8–9.

[97] See AI Act, art 51(1)(a) and recital 111, tenth sentence.

[98] See AI Act, art 51(1)(a) and recital 111, first and second sentences.

[99] See AI Act, art 51(1)(a) and recital 111, eighth sentence. Haar and Siglmüller, 'Art. 51' (n 9) para 35 define indicators as measurable variables used to quantify an AI model's performance and list accuracy, precision and recall as commonly used indicators.

'benchmarks',[100] 'criteria',[101] 'thresholds',[102] 'approximations'[103] and 'evaluations',[104] without defining them or clearly distinguishing between them – and which it may be using synonymously.[105] The precise distinction between the assessment instruments mentioned in Article 51(1)(a), as well as their differentiation from the other assessment instruments, is not entirely clear. This lack of clear delineation appears ultimately immaterial for Article 51(1)(a), however, as the terms 'technical tools' and 'methodologies' appear sufficiently broad to encompass, in principle, all conceivable types of assessment instruments for high-impact capabilities.[106] Moreover, Article 51(1)(a)'s enumeration of indicators and benchmarks as examples of technical tools and methodologies is non-exhaustive ('including'), such that thresholds[107] and evaluations for capabilities, for instance, can be subsumed under Article 51(1)(a)'s methodologies.

26. More practically significant than the precise distinction between different assessment instruments is the question of what requirements must be satisfied for such instruments to qualify as 'appropriate' within the meaning of Article 51(1)(a). Recital 111 provides an initial indication as it lays out that '[t]hresholds, as well as tools and benchmarks for the assessment of high-impact capabilities, should be *strong predictors of generality, its capabilities and associated systemic risk* of general-purpose AI models, and could take into account the way the model will be placed on the market or the number of users it may affect' (emphasis added).[108] The critical consideration appears to be whether the respective assessment instruments are sufficient predictors of the model's capabilities – and not the model's generality or its systemic risk also mentioned by the recital – as it is those capabilities that trigger classification under Article 51(1)(a).

27. Interestingly, the Safety and Security Chapter of the Code of Practice regards 'appropriate' as a less demanding standard compared to 'best practice'[109] or 'state of the art'[110], defining it as '*suitable and necessary to achieve the intended purpose* of systemic risk assessment and/or mitigation, whether through best practices, the state of the art, or other more innovative processes, measures, methodologies, methods, or techniques that go beyond the state of the art' (emphasis added).[111] The

[100] See AI Act, art 51(1)(a), annex XIII, point (e), and recital 111, eighth and tenth sentences. Haar and Siglmüller, 'Art. 51' (n 9) para 36 define benchmarks as standardised tests to measure an AI model's performance in a controlled setting; for an overview over various AI benchmarks, see Epoch AI, 'AI Benchmarking' <https://epoch.ai/benchmarks>, accessed 7 January 2026).

[101] See AI Act, annex XIII and recital 111, twelfth sentence.

[102] See AI Act, art 51(3); annex XIII, point (d); recital 111, fifth, seventh, eighth, tenth and eleventh sentences, and recital 112, fourth sentence.

[103] See AI Act, recital 111, fifth sentence.

[104] See AI Act, annex XIII, point (e).

[105] Haar and Siglmüller, 'Art. 51' (n 9) para 84.

[106] Apparently opposing view: Hofmann-Coombe (n 6) para 41 who argues that only quantitative metrics may play a role in the context of article 51(1)(a).

[107] See Koessler, Schuett and Anderljung (n 96) 8–9.

[108] AI Act, recital 111, tenth sentence; see also Haar and Siglmüller, 'Art. 51' (n 9) 32 who do not interpret recital 111's tenth sentence as providing strict requirements ('Zielvorgabe').

[109] Code of Practice Safety and Security Chapter (n 22) 27 defines 'best practice' as 'accepted amongst providers of general-purpose AI models with systemic risk as the processes, measures, methodologies, methods, and techniques that *best assess and mitigate systemic risks* at any given point in time' (emphasis added).

[110] Code of Practice Safety and Security Chapter (n 22) 30 defines 'state of the art' as 'the forefront of relevant research, governance, and technology *that goes beyond best practice*' (emphasis added).

[111] Code of Practice Safety and Security Chapter (n 22) 27.

definitions contained in the Code of Practice are not directly applicable to the AI Act but they can contribute to its literal and systematic interpretation.[112]

28. Further drawing on the Safety and Security Chapter of the Code of Practice, relevant criteria for assessing a method's appropriateness could include its scientific rigour, including its validity and reproducibility.[113] Arguably, assessment instruments that have comparable or superior predictive power to training compute thresholds should be permissible, as the legislature has deemed such thresholds a sufficient basis for Article 51(2)'s presumption of high-impact capabilities and thus appropriate. Article 51(3) empowers the Commission to introduce and specify the instruments relevant for assessing a model's high-impact capabilities.[114]

29. The AI Act does not, however, require the formal adoption of technical tools and methodologies for evaluating a model's high-impact capabilities via a delegated act.[115] Such a requirement does not find support in Article 51(1)(a)'s wording, which requires only that the assessment instruments be 'appropriate' but not that they be provided for in a delegated act. In this respect, Article 51(1)(a) differs from other provisions where the legislature expressly set out that their applicability is contingent upon the adoption of a delegated act.[116] Nor does such a requirement follow from Article 51(3). Even if this provision, by its wording ('shall'), establishes the Commission's obligation to adopt delegated acts in certain circumstances,[117] this does not support the requirement that assessment instruments must be formally adopted under Article 51(1)(a). This is because the obligation under Article 51(3) is limited to instances where the adoption of a delegated act is necessary for the thresholds under Article 51(1) and (2) to reflect the state of the art.[118]

2.1.2. Article 51(1)(b)

30. Article 51(1)(b) contains the second alternative condition for the classification of a GPAI model as presenting systemic risk. This condition comprises two main requirements. First, it requires that the

[112] See commentary on Article 56, Section 2.7.1.2. in this work for the effects of a provider's adherence to the Code of Practice.

[113] The Code of Practice Safety and Security Chapter (n 22) app 3.1 requires its signatories to 'ensure that the model evaluations are conducted with high scientific and technical rigour, ensuring: (1) internal validity; (2) external validity; and (3) reproducibility.' For a definition of these terms see Code of Practice Safety and Security Chapter (n 22) 27–28.

[114] See Section 2.3.1.

[115] The Commission Guidelines (n 39) paras 27, 31 are not entirely clear in that respect. They state that the tools and methodologies referred to under article 51(1)(a) 'are to be further specified by the Commission through adoption of delegated acts' (Commission Guidelines (n 39) para 27) and exclude the possibility that the notification obligation under article 52(1)'s first sentence could be triggered by a GPAI model having actual high-impact capabilities in absence of delegated acts providing for assessment instruments (see Commission Guidelines (n 39) para 31 and accompanying footnote). As laid out in commentary on Article 52, Section 2.1.1.2.2. in this work, this does not appear entirely convincing.

[116] See, for example, Regulation (EU) No 909/2014 of the European Parliament and of the Council of 23 July 2014 on improving securities settlement in the European Union and on central securities depositories and amending Directives 98/26/EC and 2014/65/EU and Regulation (EU) No 236/2012 [2014] OJ L 257/1, art 76(5) ('The settlement discipline measures referred to in Article 7(1) to (13) and the amendment laid down in Article 72 shall apply from the date of entry into force of the delegated act adopted by the Commission pursuant to Article 7(15).') and Regulation (EU) 2015/2365 of the European Parliament and of the Council of 25 November 2015 on transparency of securities financing transactions and of reuse and amending Regulation (EU) No 648/2012 [2015] OJ L 337/1, art 33(2)(a) ('Article 4(1) [...] shall apply: (i) 12 months after the date of entry into force of the delegated act adopted by the Commission pursuant to Article 4(9) [...]').

[117] See Section 2.3.2.

[118] AI Act, art 51(3).

GPAI model has 'capabilities or an impact equivalent to those set out in point (a)' of Article 51, that is, high-impact capabilities.[119] Annex XIII contains criteria that should be taken into account in this context.[120] Second, this determination must be 'based on a decision of the Commission, *ex officio* or following a qualified alert from the scientific panel'.[121] This requirement refers to the need for a designation decision pursuant to Article 52(4)'s first subparagraph for classification under Article 51(1)(b).[122]

31. The classification condition under Article 51(1)(b) serves 'to complement' the system of high-impact capabilities-based classification under Article 51(1)(a) by allowing the Commission 'to take individual decisions designating a general-purpose AI model as a general-purpose AI model with systemic risk' on the basis of an 'overall assessment' of Annex XIII criteria.[123] This complementary function may operate in several ways. First, whereas Article 51(1)(a) focuses on a model's capabilities,[124] Article 51(1)(b) allows for classification based on a model's impact.[125] The provision's reference to a model's impact alongside its capabilities appears to reflect the legislature's conception that 'a general-purpose AI model should be considered to present systemic risks if it has high-impact capabilities [...] *or significant impact on the internal market due to its reach*.' (emphasis added).[126] Second, to the extent that classification under Article 51(1)(a) may not account for models with particularly high capabilities in a specific domain such as offensive cyber capabilities *without high-impact capabilities across all domains*, Article 51(1)(b) could fill these gaps.[127] It has further been proposed that Article 51(1)(b) could allow the classification of a GPAI model as presenting systemic risk where its provider technically reduces the amount of computation used for the model's training to fall below Article 51(2)'s threshold without this being reflected in the model's actual capabilities.[128]

2.1.2.1. Capabilities or an impact equivalent to those set out in point (a)

32. Article 51(1)(b) allows for classification of a GPAI model as presenting systemic risk where it has 'capabilities or an impact equivalent to those set out in point (a)', meaning that classification under Article 51(1)(b) requires the model to have *capabilities or an impact equivalent to high-impact*

[119] See Section 2.1.2.1.

[120] For the role of annex XIII in the context of article 51(1)(b), see Section 2.1.2.1.2.; for an overview over annex XIII, see Section 2.4.1.

[121] AI Act, art 51(1)(b).

[122] See Section 2.1.2.2.; see also commentary on Article 52, Section 2.3.1. in this work.

[123] AI Act, recital 111, eleventh and twelfth sentences. See Janine Wendt and Domenik Wendt, *Das neue Recht der Künstlichen Intelligenz* (Nomos 2025), s 11 para 21 who argue that article 51(1)(b) allows the Commission to close gaps left by article 51(1)(a) *ad hoc* ('Ad-hoc-Lückenschließung').

[124] See AI Act, art 3(64) ('"high-impact capabilities" means capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models;') which defines 'high-impact capabilities' solely on the basis of a model's capabilities, despite the term itself referencing 'impact'.

[125] See Section 2.1.2.1.4.

[126] AI Act, recital 111, second sentence.

[127] See Section 2.1.2.1.3.1. See also Joaquin Vanschoren, 'The Role of AI Safety Benchmarks in Evaluating Systemic Risks in General-Purpose AI Models' (Publications Office of the European Union, JRC143259, 2025) <https://publications.jrc.ec.europa.eu/repository/handle/JRC143259> 7 ('A model might demonstrate sub-frontier performance on general intelligence benchmarks, yet still have advanced reasoning or knowledge acquisition, and exhibit dangerous propensities such as facilitating nuclear attacks or facilitating cyberattacks.')

[128] Schneider and Schneider, 'Art. 51' (n 4) paras 33–35. In particular, this could be relevant where the provider makes use of (knowledge) distillation (a model training technique involves a smaller "student model" and a larger "teacher model"; see Section 2.2.1.2.2.) and article 51(2)'s training compute threshold is interpreted as not accounting for the amount of computation used to train the teacher model (for an analysis of this question see Section 2.2.1.2.2.). One may note, however, that such cases may to some extent be covered by article 51(1)(a) itself, whose scope of application extends beyond article 51(2)'s compute threshold (see Section 2.1.1.).

capabilities.[129] In this context, some legal scholars have argued that both Article 51(1)(a) and Article 51(1)(b) impose equal substantive requirements, with the implication that Article 51(1)(b) – despite its different wording – requires the model to possess high-impact capabilities, just as Article 51(1)(a) does.[130] On this view, the difference between point (a) and (b) of Article 51(1) is primarily procedural in nature and would not establish substantively different thresholds for systemic risk classification under the two alternatives.[131] As will be examined below, this does not appear entirely convincing.[132] Rather than mirroring Article 51(1)(a)'s substantive requirements, Article 51(1)(b) establishes a condition for classification in its own right. The interpretation of this condition raises several questions: what is the role of Annex XIII for the assessment under Article 51(1)(b)?[133] When are a model's *capabilities* equivalent to high-impact capabilities?[134] What constitutes an *impact* equivalent to high-impact capabilities?[135] Does the *cumulative equivalence* of a model's capabilities and impact to high-impact capabilities suffice?[136] How does the *proven presence or absence of high-impact capabilities* influence classification under Article 51(1)(b)?[137] These questions are examined in the sections that follow.

33. Legal scholars have rightly noted that the criteria for classification under Article 51(1)(b) and Annex XIII leave considerable room for flexibility, such that the Commission may enjoy substantial discretion in designating a GPAI model on this basis.[138] The existence of a margin of discretion finds support in the wording of the designation provision under Article 52(4)'s first subparagraph ('may') which relates to Article 51(1)(b).[139] Moreover, Article 51(1)(b)'s wording itself suggests a substantial margin of discretion for the Commission for deciding whether the requirements for designation are met.[140] Particularly striking is the fact that the requirement of a Commission decision is placed before

[129] Commission Guidelines (n 39) para 43; interestingly, European Commission, 'General-Purpose AI Models in the AI Act – Questions & Answers' (n 4) only states that designation based on article 51(1)(b) is intended to 'capture models with *an impact* equivalent to the most advanced models' (emphasis added) without expressly mentioning equivalent capabilities.

[130] Bernsteiner and Schmitt, 'Art. 51' (n 6) para 25; see also Wendt and Wendt (n 123) s 11 para 21; Martini (n 2) para 192; Engel (n 39) 23; seemingly opposing view: Haar and Siglmüller, 'Art. 51' (n 9) paras 32, 58.

[131] Bernsteiner and Schmitt, 'Art. 51' (n 6) para 25 who argue that the central difference would be that article 51(1)(a) requires the evaluation of high-impact capabilities through appropriate technical tools and methodologies, whereas article 51(1)(b) requires evaluation according to the criteria set out in annex XIII; see also Martini (n 2) para 192.

[132] See Section 2.1.2.1.1.

[133] See Section 2.1.2.1.2.

[134] See Section 2.1.2.1.3.

[135] See Section 2.1.2.1.4.

[136] See Section 2.1.2.1.5.

[137] See Section 2.1.2.1.6.

[138] See Schneider and Schneider, 'Art. 51' (n 4) para 16 who argue that the criteria in annex XIII offer the Commission an extremely wide margin of discretion ('extrem weiten Spielraum') which may be suitable to close gaps in the classification framework but poses very limited limits to classification; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 27 emphasise that the legislature intended a flexible system in which the Commission may weight the criteria contained in annex XIII differently on a case-by-case basis and take additional criteria into account; Haar and Siglmüller, 'Art. 51' (n 9) para 59–60 refer to Article 51(1)(b) as "subjective classification" ('subjektive Einstufung') and acknowledge that the Commission enjoys very broad discretion in this context; Toby Bond and Shima Abbady, 'Article 51: Classification of General-Purpose AI Models as General-Purpose AI Models with Systemic Risk' in Ceyhun Necati Pehlivan, Nikolaus Forgó and Peggy Valcke (eds), *The EU Artificial Intelligence (AI) Act: A Commentary* (Wolters Kluwer 2024) do not expressly address the question of Commission discretion in the context of article 51(1)(b) but lament the 'significant uncertainty' around article 51(1)(b)'s interpretation.

[139] For the relationship between article 51(1)(b)'s classification condition and the designation provision under article 52(4)'s first subparagraph, see commentary on Article 52, Section 2.1.3.1. in this work; see also Section 2.1.2.2.

[140] For the difficulties of defining and delineating the relevant categories of discretionary powers under EU law and the relevance of the concrete statutory rules for determining 'the margin of manoeuvre the EU administration enjoys', see Hanns Peter Nehl, 'Judicial Review of Complex Socio-Economic, Technical, and Scientific Assessments

the substantive requirements in the provision's structure ('based on a decision of the Commission, [...] it has capabilities or an impact equivalent to those set out in point (a)'),[141] thereby creating the impression that the focus lies on the Commission's assessment rather than the objective presence of capabilities or an impact equivalent to high-impact capabilities. It may be this wording which has led some legal scholars to characterise Article 51(1)(b) as establishing 'subjective classification'.[142] Furthermore, Recital 111 refers to the Commission's 'overall assessment' of Annex XIII criteria in the context of Article 51(1)(b), without prescribing any weighting of these criteria.[143] In light of Article 51(1)(b)'s complementary purpose[144] and the legislature's concerns regarding technological developments,[145] it appears close at hand that the legislature, by creating a more flexible Article 51(1)(b) with broad Commission discretion alongside Article 51(1)(a), sought to ensure that the classification framework provided by both provisions would be future-proof and resilient to disruption.[146]

2.1.2.1.1. Distinction from Article 51(1)(a)

34. Before turning to Article 51(1)(b)'s precise requirements,[147] it is worth considering whether the substantive requirements for classification under Article 51(1)(b) are distinct from those under Article 51(1)(a).[148] Several considerations support the view that Article 51(1)(b) establishes distinct substantive requirements. First, the provisions employ different language: while Article 51(1)(a) refers to 'high-impact capabilities', Article 51(1)(b) requires 'capabilities or an impact equivalent to those set out in point (a)'.[149] Second, classification under Article 51(1)(a) is directly based *only* on

in the European Union' in Joana Mendes (ed), *EU Executive Discretion and the Limits of Law* (Oxford University Press 2019) 157, 162. For the varying terminology employed by the EU Courts with regard to discretion and margins of appreciation in the case of administrative decision-making powers under EU law, see Herwig C. H. Hofmann, 'The Interdependencies between Delegation, Discretion, and the Duty of Care' in Joana Mendes (ed), *EU Executive Discretion and the Limits of Law* (Oxford University Press 2019) 220, 223–227.

[141] This phrasing cannot be explained solely by the legislature's intention to clarify that classification under article 51(1)(b), unlike article 51(1)(a), requires Commission designation.

[142] See Haar and Siglmüller, 'Art. 51' (n 9) para 58 ('subjektive Einstufung').

[143] See AI Act, recital 111, twelfth sentence: 'That decision should be taken on the basis of an overall assessment of the criteria for the designation of a general-purpose AI model with systemic risk set out in an annex to this Regulation, such as quality or size of the training data set, number of business and end users, its input and output modalities, its level of autonomy and scalability, or the tools it has access to.'

[144] See AI Act, recital 110, tenth sentence; see also Section 2.1.2.

[145] Concerns regarding future-proofness and resilience to disruption in light of technological developments constitute a recurrent theme in the AI Act's recitals and have motivated various of its provisions (see, for example, AI Act, recital 12, first sentence; recital 101, last sentence; recital 138, second sentence; and recital 179, seventh sentence). While the recitals do not expressly mention these concerns as a basis for creating article 51(1)(b), they make clear that such considerations informed the classification rules under Section 1 of Chapter V more generally, as evidenced by recital 179's seventh sentence, stating that '[t]he AI Office should ensure that classification rules and procedures are up to date in light of technological developments', and recital 111's sixth sentence, which by its reference to 'the state of the art at the time of entry into force of this Regulation' implicitly acknowledges that the best approximations for model capabilities may shift over time.

[146] See also AI Act, recital 97, thirteenth sentence: 'Considering their potential significantly negative effects, the general-purpose AI models with systemic risk should always be subject to the relevant obligations under this Regulation.'

[147] See Section 2.1.2.1.3. ('Equivalent capabilities'); Section 2.1.2.1.4. ('Equivalent impact'); Section 2.1.2.1.5. ('Cumulative equivalence').

[148] In favour: Bond and Abbady, 'Art. 51' (n 138) 833, s 3.2; Hofmann-Coombe (n 6) para 48; against: Bernsteiner and Schmitt, 'Art. 51' (n 6) para 25; Wendt and Wendt (n 123) s 11 para 21; Martini (n 2) para 192; Engel (n 39) 23.

[149] See Bond and Abbady, 'Art. 51' (n 138) 833, s 3.2 ('Note that Article 51(1)(b) adds 'impact' as a relevant criterion, where this is not mentioned in Article 51(1)(a).')

the model's high-impact capabilities, including indicators and benchmarks for such capabilities, whereas Article 51(1)(b) requires the Commission to have regard to the criteria in Annex XIII,[150] some of which concern the model's reach[151] – such as the number of registered business and end users (points (f) and (g)) – rather than its capabilities. These criteria serve as indicators for the model's impact, and do not indicate, or indicate only tangentially, whether a model has high-impact capabilities.[152]

35. Third, training compute plays a fundamentally different role under each condition. Under Article 51(1)(a), exceeding the training compute threshold in Article 51(2) alone creates a presumption of high-impact capabilities[153] and is therefore sufficient for classification.[154] Under Article 51(1)(b), however, the amount of training compute is merely one criterion among several listed in Annex XIII that inform the model's classification,[155] an approach which likely aims to enable classification of models staying below the Article 51(2) threshold.[156] Fourth, Article 52 contains two distinct provisions for designation of GPAI models as presenting systemic risk,[157] one arguably permitting designation on the basis of Article 51(1)(a) and the other on the basis of Article 51(1)(b).[158] If both conditions applied the same substantive standard, the existence of two separate designation procedures would be difficult to explain. Finally, if Article 51(1)(b) merely replicated Article 51(1)(a)'s substantive requirements, it would risk being redundant:[159] models meeting those requirements would already automatically be classified under Article 51(1)(a).[160]

36. An interpretation of Article 51(1)(b) as permitting classification without requiring high-impact capabilities can be reconciled with Article 3(65)'s definition of systemic risk.[161] If systemic risk refers to risk that is '*specific to* the high-impact capabilities of general-purpose AI models' (emphasis added), the question arises whether the classification of GPAI models without high-impact capabilities as presenting systemic risk can be justified. Such classification is possible. As discussed below, it is uncertain whether Article 3(65) necessarily implies that systemic risk can only arise in GPAI models with high-impact capabilities or whether it is merely characteristic of models with

[150] See Section 2.1.2.1.2.

[151] A model's reach is one of the relevant indicators for its impact (see AI Act, art 3(65)); see also Section 2.1.2.1.4.

[152] See Bond and Abbady, 'Art. 51' (n 138) 833, s 3.2 ('However, including criteria related to widespread use in the assessment of systemic risk does not appear to be entirely consistent with Article 51(1)(a), which determines that GPAI models (which are not designated by the Commission ex officio) qualify as systemic risk models if they have "high-impact *capabilities*". That is unless one views widespread use as an indicator of capability, which will not always be obvious in practice. It appears rather that widespread use in Annex XIII is viewed as an indicator of "impact", as Article 51(1)(b) and Annex XIII both provide that models which are in scope are those that have a certain level of "capabilities or impact"'.)

[153] See Section 2.2.2.

[154] See Section 2.2.2. However, the provider may challenge its model's classification pursuant to article 52(2) and (3).

[155] See AI Act, annex XIII, point (c).

[156] See Schneider and Schneider, 'Art. 51' (n 4) paras 33–35.

[157] AI Act, art 52(1), third sentence, and art 52(4), first subparagraph.

[158] See commentary on Article 52, Section 2.1.3.1. in this work.

[159] For the principle that 'where a provision of EU law is open to several interpretations, preference must be given to that interpretation which ensures that the provision retains its effectiveness' see, for example, Case C-154/21 *RW v Österreichische Post AG* [2023] ECLI:EU:C:2023:3 para 29 and the case law cited therein.

[160] See Section 2.1.1.1.

[161] According to article 3(65), systemic risk is 'a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain'.

such capabilities.[162] But even in the former case, the proven absence of high-impact capabilities is not the same as *uncertainty* about their presence. Accordingly, the fact that Article 51(1)(b) does not require the Commission to establish the model's high-impact capabilities for classification does not necessarily mean it permits designation of models proven not to possess such capabilities.[163] Rather, it permits the Commission to designate a GPAI model according to a distinct substantive standard – capabilities or an impact equivalent to high-impact capabilities – even in cases where evidence concerning the presence of high-impact capabilities is inconclusive. After a minimum period of six months, a provider can request reassessment of its model's designation on the basis of new reasons that have arisen since the designation decision.[164]

37. As indicated above, some legal scholars nevertheless contend that the conditions in point (a) and (b) of Article 51(1) do not entail different substantive requirements for classification.[165] They argue that Article 51(1)(b)'s express reference to Article 51(1)(a) indicates substantive equivalence, and that the difference between these conditions can be explained on procedural grounds.[166] According to this view, the distinction between Article 51(1)(a)'s evaluation of the model's capabilities on the basis of appropriate technical tools and methodologies and Article 51(1)(b)'s determination with regard to the criteria set out in Annex XIII merely reflects the provider's duty to conduct an actual evaluation of its model.[167] This procedural explanation is not entirely convincing,[168] as procedural reasons may explain Article 51(1)(a)'s reference of 'appropriate technical tools and methodologies' and Article 51(1)(b)'s reference of Annex XIII but cannot explain why Article 51(1)(b) requires 'capabilities or an impact equivalent to those set out in point (a)' rather than simply 'high-impact capabilities'.

2.1.2.1.2. Role of Annex XIII criteria

38. Annex XIII contains criteria[169] which are particularly relevant for determining whether a model has capabilities or an impact equivalent to high-impact capabilities.[170] The importance of these criteria for this assessment under Article 51(1)(b) is evidenced by the provision's express reference to Annex XIII and reinforced by the introductory sentence of this annex, which expressly states that it contains a list of criteria that the Commission shall take into account '[f]or the purpose of determining that a general-purpose AI model has capabilities or an impact equivalent to those set out in Article 51(1), point (a)'.

---

[162] Under a literal reading, 'specific to' can either mean *exclusive to* – implying that only GPAI models with high-impact capabilities can present systemic risks – or *characteristic of* – implying that GPAI models with high-impact capabilities typically present systemic risks without excluding that GPAI models without such capabilities may under certain circumstances present systemic risks as well; see the discussion Section 2.1.2.1.6.2.

[163] For a discussion on whether the Commission can designate a GPAI model as presenting systemic risk where it does not have high-impact capabilities see Section 2.1.2.1.6.2.

[164] AI Act, art 52(5). See commentary on Article 52, Section 2.4. in this work.

[165] Bernsteiner and Schmitt, 'Art. 51' (n 6) para 25; see also Wendt and Wendt (n 123) s 11 para 21; Martini (n 2) para 192; Engel (n 39) 23; seemingly opposing view: Haar and Siglmüller, 'Art. 51' (n 9) paras 32, 58.

[166] Bernsteiner and Schmitt, 'Art. 51' (n 6) para 25.

[167] Bernsteiner and Schmitt, 'Art. 51' (n 6) para 25 ('Dass die Fähigkeiten des KI-Modells einmal mithilfe geeigneter technischer Instrumente und Methoden (lit. a), einmal unter Berücksichtigung der in Anh. XIII festgelegten Kriterien (lit. b) ermittelt werden sollen, ist darauf zurückzuführen, dass der Anbieter eine tatsächliche Prüfung und Bewertung seines Modells durchführen und im Zuge dessen auch die Risikogeneigtheit bewerten soll.').

[168] Similar: Haar and Siglmüller, 'Art. 51' (n 9) para 32.

[169] For an overview over these criteria, see Section 2.4.1.

[170] See Bond and Abbady, 'Art. 51' (n 138) 835, s 4.1; Haar and Siglmüller, 'Art. 51' (n 9) para 58; European Commission, 'General-Purpose AI Models in the AI Act – Questions & Answers' (n 4).

39. The subsequent sections discuss whether and to what extent the Commission is required to take the criteria contained in this annex into account when designating a GPAI model as presenting systemic risk on the basis of Article 51(1)(b) in conjunction with Article 52(4)'s first subparagraph,[171] and argue that the list of criteria contained in Annex XIII is non-exhaustive in this context.[172]

2.1.2.1.2.1. Mandatory consideration

40. Multiple provisions suggest a duty for the Commission to consider the criteria contained in Annex XIII when designating a GPAI model as presenting systemic risk on the basis of Article 51(1)(b) in conjunction with Article 52(4)'s first subparagraph.[173] Annex XIII itself establishes that the Commission 'shall take into account [its] criteria' for the purpose of determining that a GPAI model has capabilities or an impact equivalent to those set out in Article 51(1)(a).[174] Article 51(1)(b) reinforces this by requiring the Commission to '[have] regard to the criteria set out in Annex XIII'. Article 52(4)'s first subparagraph and Recital 111 further confirm that this decision must be taken 'on the basis of criteria set out in Annex XIII' and 'on the basis of an *overall assessment* of [those] criteria' (emphasis added), respectively. The Commission's duty to consider the Annex XIII criteria is reminiscent of the Commission's general duty of diligent and impartial examination, which requires it to carefully examine the relevant facts of a case[175] and is – at least for some areas of EU law such as competition law and state aid – recognised by the EU Courts.[176]

41. This raises the question of how this duty relates to the substantive requirements for designation under Article 51(1)(b), in particular whether the Commission must enquire about *all* of Annex XIII's criteria to determine that a GPAI model has capabilities or an impact equivalent to high-impact capabilities or whether it can base its designation decision on only some of the criteria.[177] Compelling arguments can be made in support of both interpretations. In favour of the Commission's discretion to selectively focus on some criteria, one might argue that Article 51(1)(b)'s use of 'or' permits the Commission to base its designation decision either on the model's impact or on the model's capabilities and therefore to focus solely on the Annex XIII criteria relating to capabilities while ignoring those concerning impact, or vice versa.[178] For example, where the Commission comes to the conclusion that a GPAI model has capabilities equivalent to high-impact

[171] See Section 2.1.2.1.2.1.
[172] See Section 2.1.2.1.2.2.
[173] The Commission decision referred to under Article 51(1)(b) arguably constitutes a designation decision in the sense of article 52(4)'s first subparagraph (see Section 2.1.2.2.).
[174] Apparently unconvinced by this textual argument: Bond and Abbady, 'Art. 51' (n 138) 834, s 3.2 ('[I]t is not clear from the text whether the criteria in Annex XIII are cumulative and/or exhaustive.')
[175] See Paul Craig, *EU Administrative Law* (Oxford University Press, 3rd edn, 2018) ch 12 s 3(A).
[176] For the application of this duty in administrative procedures entailing complex technical evaluations see Case C-269/90 *Hauptzollamt München-Mitte v Technische Universität München* [1991] ECR I-5469 paras 13–14 ('It must be stated first of all that, since an administrative procedure entailing complex technical evaluations is involved, the Commission must have a power of appraisal in order to be able to fulfil its tasks. However, where the Community institutions have such a power of appraisal, respect for the rights guaranteed by the Community legal order in administrative procedures is of even more fundamental importance. Those guarantees include, in particular, the duty of the competent institution to examine carefully and impartially all the relevant aspects of the individual case, the right of the person concerned to make his views known and to have an adequately reasoned decision.'); for the duty of diligent and impartial examination with respect to state aid examinations Case C-59/24 P *Kingdom of the Netherlands v European Commission* [2025] ECLI:EU:C:2025:798 para 88; for a general discussion of this duty and its relevance in EU administrative law see Craig (n 175) ch 12, s 3.
[177] See Bond and Abbady, 'Art. 51' (n 138) 834, s 3.2 ('[I]t is not clear from the text whether the criteria in Annex XIII are cumulative and/or exhaustive.')
[178] For the meaning of 'or' in article 51(1)(b), see also Section 2.1.2.1.5.

capabilities based on Annex XIII's capabilities-related criteria (points (a)–(e)), one may argue it would not also have to enquire about the model's number of registered business and end users (Annex XIII, points (f) and (g)). This would appear sensible, insofar as the number of registered users appears to be a rather irrelevant approximation for a model's capabilities in many cases.[179] However, this selective approach sits in tension with the abovementioned provisions signalling a duty to take Annex XIII criteria into account. Given that Annex XIII establishes that the Commission 'shall' rather than 'may' take the criteria into account and Recital 111 refers to an 'overall assessment',[180] this implies mandatory consideration of all Annex XIII criteria.[181]

42. Two arguments nevertheless support a flexible interpretation of the Commission's duty to consider Annex XIII criteria, such that the Commission enjoys discretion in its decision over which Annex XIII criteria to focus on[182] – just as Article 51(1)(b) allows the Commission to consider either capabilities or impact.[183] First, the wording of 'take into account',[184] 'having regard to'[185] and 'on the basis of'[186] does not appear to impose a particularly stringent requirement. In particular, it does not imply that any single criterion necessarily qualifies or disqualifies a model with regard to Article 51(1)(b)'s standard.[187] Rather, it appears sufficient to engage in a reasoned analysis demonstrating that a particular criterion does not, in the individual case, contribute to the decision on whether Article 51(1)(b) is met, thereby having regard to that criterion. Second, an overly strict interpretation of the Commission's duty to consider Annex XIII criteria risks creating perverse incentives for providers. Such an interpretation could encourage providers to withhold information – such as the precise number of parameters – when responding to requests from the AI Office for the model's technical documentation[188] or other information[189], thereby preventing designation of their model through non-cooperation. Nevertheless, where information regarding a relevant criterion is lacking, the Commission will generally be obliged to afford the provider an opportunity to supply it, as follows from the provider's right to be heard before a designation decision is made.[190]

[179] See Bond and Abbady, 'Art. 51' (n 138) 833, s 3.2.

[180] See AI Act, annex XIII, introductory sentence, and recital 111, twelfth sentence.

[181] See Hofmann-Coombe (n 6) para 49; Schneider and Schneider, 'Art. 51' (n 4) para 15; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 27; Haar and Siglmüller, 'Art. 51' (n 9) para 60.

[182] See Schneider and Schneider, 'Art. 51' (n 4) para 16 who argues that annex XIII offers the Commission a wide margin of discretion and that, while its criteria need to be taken into account, none of them necessarily qualifies or disqualifies a model; see also Bernsteiner and Schmitt, 'Art. 51' (n 6) para 27; unclear: Samuel Carey, 'Regulating Uncertainty: Governing General-Purpose AI Models and Systemic Risk', (2025) European Journal of Risk Regulation <https://doi.org/10.1017/err.2025.10040>, 9 ('It is unclear whether these benchmarks and indicators are to be interpreted exhaustively or selectively, and if selectively, it does not provide any instruction as to how each indicator should be weighed against the other.')

[183] AI Act, art 51(1)(b); see also Section 2.1.2.1.5.

[184] AI Act, annex XIII.

[185] AI Act, art 51(1)(a).

[186] AI Act, art 52(4), first subparagraph, and recital 111, twelfth sentence.

[187] Schneider and Schneider, 'Art. 51' (n 4) para 16.

[188] See AI Act, art 53(1)(a).

[189] See AI Act, art 91.

[190] For this right to prior hearing, see AI Act, art 94 in conjunction with Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011, OJ L 169/1 ("MSR"), art 18(3); see also commentary on Article 52, Section 2.1.3.2. in this work.

2.1.2.1.2.2. Non-exhaustive nature

43. A further question concerns whether Annex XIII is exhaustive – that is, whether the Commission may also take into account additional criteria beyond those listed, such as the model's architecture, training methods, or foreseeable negative effects, when designating a model under Article 51(1)(b) in conjunction with Article 52(4)'s first subparagraph.[191] The more compelling view is that Annex XIII is non-exhaustive.[192] In the absence of other conclusive textual indicators,[193] both Article 51(1)(b)'s wording ('having regard to') and Annex XIII's wording ('take into account') suggest that other criteria may be taken into account for designation under Article 51(1)(b) in conjunction with Article 52(4), first subparagraph.[194] Moreover, the inclusion of broad and open-ended criteria in Annex XIII – with non-exhaustive lists of examples[195] – suggests that the legislature sought to ensure consideration of diverse factors rather than to confine the Commission to a closed list.[196].

44. This interpretation finds further support in the fact that relevant criteria are not mentioned in the list. In particular, while Annex XIII expressly mentions the number of parameters, it does not mention model architecture.[197] This does not appear to be a conscious omission, as both the number of parameters and model architecture are listed together in point 1(d) of Section 1 of Annex XI as part of the technical documentation referred to in Article 53(1)(a) which providers must, upon request, provide to the AI Office and national competent authorities.[198] Moreover, information about the model's architecture provides helpful context for information about the number of parameters.[199] Further, Recital 111's tenth sentence mentions 'the way the model will be placed on the market' as a relevant criterion for systemic risk classification, which is not expressly reflected in Annex XIII itself. Similarly, while Article 3(65) defines systemic risk by reference to both (i) the model's reach and (ii) its 'actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or society as a whole' as indicators of the model's impact, the Annex expressly mentions only the former and not the latter, even though both appear side by side

---

[191] See Bond and Abbady, 'Art. 51' (n 138) 834, s 3.2 ('[I]t is not clear from the text whether the criteria in Annex XIII are cumulative and/or exhaustive.')

[192] Bernsteiner and Schmitt, 'Art. 51' (n 6) 27.

[193] Bond and Abbady, 'Art. 51' (n 138) 834, s 3.2. Article 51(1)(b), article 52(4)'s first paragraph and annex XIII lack common indicators that signal whether a list is exhaustive or non-exhaustive for a specific purpose. A typical indicator of an non-exhaustive list under EU law is the phrase 'inter alia' (see, for example, Directive 2011/95/EU of the European Parliament and of the Council of 13 December 2011 on standards for the qualification of third-country nationals or stateless persons as beneficiaries of international protection, for a uniform status for refugees or for persons eligible for subsidiary protection, and for the content of the protection granted (recast) [2011] OJ L 337/9, art 9(2)). A typical indicator of an exhaustive list under EU law is the word 'only' (see, for example, Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora [1992] OJ L 206/7, art 6(4)).

[194] Bernsteiner and Schmitt, 'Art. 51' (n 6) para 27. For the use of 'having regard to' in an apparently non-exhaustive way see, for example, Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006 [200] OJ L 353/1, art 54(2): 'Where reference is made to this paragraph, Articles 5 and 7 of Decision 1999/468/EC shall apply, having regard to the provisions of Article 8 thereof.'

[195] See AI Act, art 53(1)(a) and annex XIII, point (d) and (e).

[196] See Bernsteiner and Schmitt, 'Art. 51' (n 6) 27.

[197] AI Act, annex XIII, point (a).

[198] See commentary on Article 53, paras 9–10 in this work.

[199] For example, so-called mixture-of-expert models have a smaller number of *active* parameters than traditional dense models do, which is relevant when comparing the total number of parameters of those models (see Ege Erdil, 'How Do Mixture-Of-Experts Models Compare to Dense Models in Inference?' (2024) <https://epoch.ai/gradient-updates/moe-vs-dense-models-inference> accessed 7 January 2026).

in that definitional provision. While it might be possible and necessary to interpret Annex XIII criteria extensively to encompass considerations not expressly stated,[200] these omissions further suggest that Annex XIII is non-exhaustive.

45. The question of Annex XIII's exhaustiveness is connected to the question of the scope of the Commission's power to amend Annex XIII under Article 52(4)'s second subparagraph, specifically whether the Commission is permitted to add criteria to Annex XIII.[201] As laid out elsewhere, there are compelling reasons for interpreting this delegation of power broadly as allowing the Commission to add new criteria.[202]

2.1.2.1.3. Equivalent capabilities

46. The capabilities of a GPAI model generally refer to its ability to perform various tasks.[203] Points (a)–(e) of Annex XIII contain criteria particularly relevant for assessing model's capabilities,[204] namely: (a) its number of parameters; (b) the quality and size of its (training) data set; (c) the amount of computation used for its training; (d) its input and output modalities; and (e) its performance on capability benchmarks and evaluations.[205] Moreover, the Safety and Security Chapter of the Code of Practice lists capabilities which may contribute to a model presenting systemic risk, including offensive cyber capabilities, capabilities to adaptively learn new tasks or capabilities to evade human oversight.[206]

47. The AI Act does not expressly specify when a model's capabilities are equivalent to high-impact capabilities. In line with Article 51(1)(b)'s interpretation as distinct from Article 51(1)(a),[207] it is not convincing to equate 'capabilities [...] equivalent to high-impact capabilities' under Article 51(1)(b) with 'high impact capabilities' under Article 51(1)(a), as Article 51(1)(b) requires 'equivalent' and not 'the same' capabilities and such an interpretation would render the inclusion of 'capabilities' alongside 'impact' in Article 51(1)(b) redundant.[208] Given that the standards diverge, the key question becomes how they differ. Any specification of 'capabilities [...] equivalent to [high-impact capabilities]' necessarily presupposes a prior determination of what qualifies as a high-impact capability.[209]

---

[200] See, for example, the discussion below (Section 2.1.2.1.4.) for the relevance of a model's actual or reasonably foreseeable negative effects for impact-based classification under article 51(1)(b).

[201] See commentary on Article 52, Section 2.3.2. in this work.

[202] See commentary on Article 52, Section 2.3.2. in this work.

[203] See the definition of a GPAI model under article 3(63) as 'capable of performing a wide range of distinct tasks' as well as the corresponding first sentence of Recital 97 which states that '[t]he definition [of a general-purpose AI model] should be based on the key functional characteristics of a general-purpose AI model, in particular the generality and the capability to competently perform a wide range of distinct tasks.' For an in-depth discussion of the concept of capability see forthcoming commentary on Article 3(63) in this work.

[204] See Bond and Abbady, 'Art. 52' (n 70) 833–834, s 3.2.

[205] For the role of annex XIII in the context of article 51(1)(b), see Section 2.1.2.1.2.; for a general overview over annex XII, see Section 2.4.1.

[206] Code of Practice Safety and Security Chapter (n 22) app 1.3.1; see also AI Act, recital 110, third sentence; for the role of codes of practice with regard to GPAI model regulation under the AI Act, see commentary on Article 56, Section 1.1. in this work.

[207] See Section 2.1.2.1.1.

[208] See Hofmann-Coombe (n 6) para 48.

[209] This question is particularly relevant for capabilities-based classification under article 51(1)(b), as classification under article 51(1)(a) is capabilities-based as well. For impact-based classification under article 51(1)(b), see Section 2.1.2.1.4.

48. This question matters for two interconnected reasons. First, the notion of high-impact capabilities forms part of the standard of capabilities 'equivalent to those set out in point (a)' under Article 51(1)(b), thus directly determining the provision's scope. Second, high-impact capabilities do not only form part of Article 51(1)(b)'s standard but they themselves also form the basis for automatic classification under Article 51(1)(a) and arguably constitute grounds for designation under Article 52(1), third sentence.[210] Understanding when a model has high-impact capabilities is therefore essential to determine the scope of classification under these provisions, and by extension how Article 51(1)(b)'s classification of GPAI models with capabilities equivalent to high-impact capabilities could meaningfully complement these provisions.[211]

49. As laid out above, the definition of high-impact capabilities under Article 3(64) raises various interpretive issues:[212] for example, what constitutes the 'most advanced' GPAI models – is advancement determined in a domain-specific assessment (such as in coding or video generation) or by overall advancement, and if the latter, how is overall advancement assessed? Once the most advanced models are identified, which capabilities should be compared – must a model 'match or exceed' all capabilities of these models, or only relevant ones, and which capabilities would be relevant? Does Article 3(64)'s reference to 'recorded' capabilities further limit the capabilities under consideration? The resolution of these interpretive questions – particularly if narrower interpretations of the notion of high-impact capabilities prevail – could give rise to different ways in which Article 51(1)(b)'s classification of models with capabilities equivalent to high-impact capabilities could complement Article 51(1)(a). In principle, a narrow interpretation of Article 51(1)(a) correspondingly expands Article 51(1)(b)'s scope to capture GPAI models that may present systemic risks yet are not classified under Article 51(1)(a).

2.1.2.1.3.1. Sufficiency of a model's domain-specific high-impact capabilities

50. Article 51(1)(b) arguably allows for the classification of GPAI models on the basis of *domain-specific* high-impact capabilities – understood here as high-impact capabilities relating to specific environments into which a GPAI model may be deployed – including through integration into an AI system[213] – which are characterised in particular by their relevant modalities and tasks.[214] This could include classification of models with particularly high chemical, biological, radiological, and nuclear ("CBRN") or offensive cyber capabilities. Such classification assumes particular significance

[210] For automatic classification under article 51(1)(a), see Section 2.1.1.1. For the requirements for designation under article 52(1), third sentence, see commentary on Article 52, Section 2.1.3.3. in this work.

[211] For article 51(1)(b)'s complementary purpose see Section 2.1.2. Moreover, it is a general principle of EU law interpretation that a provision should not be interpreted in a manner that renders it redundant (see Cases *RW v Österreichische Post AG* (n 159) para 29 and C-31/17 *Cristal Union, the legal successor to Sucrerie de Toury SA v Ministre de l'Économie et des Finances* [2018] ECLI:EU:C:2018:168 para 41; Koen Lenaerts and José A. Gutiérrez-Fons, 'To Say What the Law of the EU Is: Methods of Interpretation and the European Court of Justice' (2014) 20 Columbia Journal of European Law 3, 17–21).

[212] See Section 2.1.1.2.; see also forthcoming commentary on Article 3(64) in this work.

[213] See AI Act, recital 97, eighth sentence: 'AI models are typically integrated into and form part of AI systems.'

[214] See Yoshua Bengio and others, 'International AI Safety Report' (DSIT 2025/001, 2025) <https://internationalaisafetyreport.org/publication/international-ai-safety-report-2025> 224 which defines '[o]pen-ended domains' as '[e]*nvironments into which AI systems might be deployed* which present a very large set of possible scenarios.', (emphasis added). This definition encompasses domains such as science, software engineering, mathematics, health care, structural biology, planning, game-playing, natural language processing, computer vision, speech recognition and image classification – in line with the use of the term in Bengio and others 17, 24, 27, 48, 51–52, 57, 58, 111, 159.

if Article 51(1)(a) only encompasses cross-domain high-impact capabilities – a reading which remains uncertain but appears plausible given Article 3(64)'s wording.[215]

51. Annex XIII supports classification under Article 51(1)(b) on the basis of *domain-specific* high-impact capabilities. While Annex XIII could in principle be used to assess both overall high-impact capabilities and domain-specific high-impact capabilities, two features of this annex indicate that the legislature drafted it with domain-specific assessments of high-impact capabilities in mind. First, point (d) of Annex XIII refers to the '*specific* type of inputs and outputs (e.g. biological sequences)' (emphasis added) as a relevant criterion distinct from the broader criterion of 'input and output modalities of the model'. If only cross-domain capabilities mattered, the mentioning of specific types of input and output would be redundant, as all modalities would necessarily be relevant – a point already conveyed by the broader criterion.[216] This suggests that specific types of input and output and the corresponding domain-specific capabilities to process them matter in themselves. Moreover, the example of 'biological sequences' appears specifically tailored to a domain-specific capability assessment, namely that of biological capabilities.[217] Second, point (d) of Annex XIII refers to 'state of the art *thresholds* for determining high-impact capabilities *for each modality*' (emphasis added) as a further relevant criterion. Meeting a modality-specific threshold would carry little significance in itself if only a model's cross-domain high-impact capabilities mattered.[218] In addition, given that point (e) of Annex XIII mentions 'benchmarks and evaluations of capabilities' separately from the modality-specific thresholds in point (d), this distinction further indicates that domain-specific capabilities are relevant in themselves rather than serving merely as indicators of a model's overall capabilities.

52. Moreover, domain-specific capabilities such as CBRN or offensive cyber capabilities[219] appear to be particularly relevant to certain systemic risks a GPAI model may cause, such as the risks of biological or cyber-attacks.[220] Were these not covered by either Article 51(1)(a) or Article 51(1)(b), some GPAI models that present such risks would fall outside Article 51(1)'s classification framework and corresponding systemic risks would be left unaddressed. However, that the legislature was aware of such domain-specific capabilities and intended them to be covered finds support in point (d) of Annex XIII mention of 'biological sequences', an example that fits a domain-specific assessment of biological capabilities.[221]

---

[215] Article 3(64) does not expressly distinguish between different domains but only refers to 'the most advanced general-purpose AI models'. For the interpretive uncertainty surrounding the article 3(64)'s definition of high-impact capabilities see Bond and Abbady, 'Art. 51' (n 138) 831 s 3.2 ('"the most advanced" is a highly open-ended concept').

[216] See AI Act, annex XIII, point (d).

[217] See AI Act, recital 110, third sentence; see also Code of Practice Safety and Security Chapter (n 22) app 1.3.1, point (2).

[218] The assessment of a model's cross-domain high-impact capabilities would require assessing all relevant modalities that matter for these domains.

[219] For the nature of these capabilities as potential sources of systemic risks see AI Act, recital 110, third sentence; Code of Practice Safety and Security Chapter (n 22) app 1.3.1, points (1) and (2).

[220] For the nature of these risks as systemic risks see AI Act, recital 110 and Code of Practice Safety and Security Chapter (n 22) app 1.4, points (1) and (2) which list these risks as specified systemic risks.

[221] See also AI Act, recital 110 mentioning 'chemical, biological, radiological, and nuclear risks' and 'offensive cyber capabilities'; AI Act, recital 97: 'Considering their potential significantly negative effects, the general-purpose AI models with systemic risk should always be subject to the relevant obligations under this Regulation.'

2.1.2.1.3.2. Interpretation of a model's domain-specific high-impact capabilities

53. On the basis that Article 51(1)(b) allows for a model's classification solely based on domain-specific high-impact capabilities, further questions arise as to which domains are relevant and how one determines high-impact capabilities in these domains. Given Article 51(1)(b)'s integration into the context of systemic risk classification, it appears appropriate to focus on capabilities that are particularly relevant to the systemic risks that GPAI models may pose.[222] In that regard, the fourteen model capabilities listed in the Safety and Security Chapter of the Code of Practice as potential sources of systemic risk merit particular consideration.[223] These include not only the aforementioned offensive cyber and biological capabilities but also further capabilities such as the capacity for manipulation, autonomous operation, and the evasion of human oversight.[224] Further regard should be had to how the AI Act addresses domain-specific capabilities. Some capabilities are mentioned expressly, such as 'offensive cyber capabilities'.[225] Others are addressed indirectly: biological capabilities, for instance, are referenced through Recital 110's mention of biological risks and Annex XIII's reference to biological sequences as model input. Moreover, international approaches that identify certain capabilities as particularly relevant to systemic risk could also be taken into account.[226]

54. The determination of whether a GPAI model has high-impact capabilities in a specific domain requires appropriate assessment instruments, just as does the determination under Article

---

[222] Recital 110 contains a non-exhaustive list of systemic risks that GPAI models could pose, stating that '[g]eneral-purpose AI models could pose systemic risks which include, but are not limited to, any actual or reasonably foreseeable negative effects in relation to major accidents, disruptions of critical sectors and serious consequences to public health and safety; any actual or reasonably foreseeable negative effects on democratic processes, public and economic security; the dissemination of illegal, false, or discriminatory content. [...] In particular, international approaches have so far identified the need to pay attention to risks from potential intentional misuse or unintended issues of control relating to alignment with human intent; chemical, biological, radiological, and nuclear risks, such as the ways in which barriers to entry can be lowered, including for weapons development, design acquisition, or use; offensive cyber capabilities, such as the ways in vulnerability discovery, exploitation, or operational use can be enabled; the effects of interaction and tool use, including for example the capacity to control physical systems and interfere with critical infrastructure; risks from models of making copies of themselves or "self-replicating" or training other models; the ways in which models can give rise to harmful bias and discrimination with risks to individuals, communities or societies; the facilitation of disinformation or harming privacy with threats to democratic values and human rights; risk that a particular event could lead to a chain reaction with considerable negative effects that could affect up to an entire city, an entire domain activity or an entire community.'; see also forthcoming commentary on Article 3(65) in this work.

[223] Code of Practice Safety and Security Chapter (n 22) app 1.3.1 lists '(1) offensive cyber capabilities; (2) Chemical, Biological, Radiological, and Nuclear (CBRN) capabilities, and other such weapon acquisition or proliferation capabilities; (3) capabilities that could cause the persistent and serious infringement of fundamental rights; (4) capabilities to manipulate, persuade, or deceive; (5) capabilities to operate autonomously; (6) capabilities to adaptively learn new tasks; (7) capabilities of long-horizon planning, forecasting, or strategising; (8) capabilities of self-reasoning (e.g. a model's ability to reason about itself, its implementation, or environment, its ability to know if it is being evaluated); (9) capabilities to evade human oversight; (10) capabilities to self-replicate, self-improve, or modify its own implementation environment; (11) capabilities to automate AI research and development; (12) capabilities to process multiple modalities (e.g. text, images, audio, video, and further modalities); (13) capabilities to use tools, including "computer use" (e.g. interacting with hardware or software that is not part of the model itself, application interfaces, and user interfaces); and (14) capabilities to control physical systems.'; for the role of codes of practice with regard to GPAI model regulation under the AI Act, see commentary on Article 56, Section 1.1. in this work.

[224] See Code of Practice Safety and Security Chapter (n 22) app 1.3.1, points (4), (5) and (9).

[225] AI Act, recital 110.

[226] See AI Act, recital 110 which refers to 'international approaches' with regard to relevant systemic risks requiring consideration.

51(1)(a).[227] This likely includes the use of domain-specific capabilities benchmarks such as the Biological Laboratory Protocol Benchmark (BioLP-bench)[228] and the Language Agent Biology Benchmark (LAB-Bench)[229] for AI models' biological capabilities.[230] Article 3(64)'s standard of 'capabilities recorded in the most advanced general-purpose AI models' can guide the determination of which domain-specific capabilities are sufficient to satisfy Article 51(1)(b)'s standard of 'capabilities [...] equivalent to [high-impact capabilities]' as the most advanced models can be determined both across domains and domain-specifically.[231]

2.1.2.1.3.3. Further interpretations of capabilities equivalent to high-impact capabilities

55. Beyond allowing for the classification of models with domain-specific high-impact capabilities, Article 51(1)(b)'s reference to capabilities equivalent to high-impact capabilities could complement classification under Article 51(1)(a) in further ways.[232] While a detailed analysis of Article 3(64)'s definition of high-impact capabilities – which would be necessary to examine all possible ways in which Article 51(1)(b) could complement Article 51(1)(a) – exceeds the scope of this chapter,[233] one possibility appears worth discussing: whether Article 51(1)(b) establishes a *lower capabilities threshold* than Article 51(1)(a) does, such that capabilities insufficient for automatic classification could nonetheless suffice for designation standing alone, without additional offsetting factors such as the model's high impact.[234]

56. Such an interpretation, while ultimately less convincing, finds some support in the fact that classification under Article 51(1)(b), in contrast to classification under Article 51(1)(a),[235] requires a Commission designation decision,[236] a requirement which arguably compensates for a lower substantive threshold by providing legal certainty with regard to a model's classification. It would read 'equivalent' as a deliberate choice of wording over 'equal', thus permitting a potentially different (lower) level of capabilities.

57. However, several considerations weigh against this interpretation. While the wording of Article 51(1)(b) does not preclude it, neither does it provide strong support, as 'equivalent' does not necessarily mean that lesser capabilities suffice. More importantly, treating capabilities below Article 51(1)(a)'s threshold as sufficient for Article 51(1)(b) appears unjustified given the substantively

---

[227] See AI Act, art 51(1)(a): 'high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks'. For a general discussion of domain-specific compute thresholds, see Lennart Heim and Leonie Koessler, 'Training Compute Thresholds: Features and Functions in AI Regulation' (2024) <https://arxiv.org/abs/2405.10799> accessed 7 January 2026 20–21 For an analysis of the requirement of appropriate assessment instruments under article 51(1)(a), see Section 2.1.1.3.

[228] Igor Ivanov, 'BioLP-bench: Measuring Understanding of Biological Lab Protocols by Large Language Models' (2024) <https://www.biorxiv.org/content/10.1101/2024.08.21.608694v3> accessed 7 January 2026.

[229] Jon M Laurent and others, 'LAB-Bench: Measuring Capabilities of Language Models for Biology Research' (2024) <https://arxiv.org/abs/2407.10362> accessed 7 January 2026.

[230] The legislature appears to presuppose the use of domain-specific benchmarks in point (d) of annex XIII by referencing 'state of the art thresholds for determining high-impact capabilities for each modality, and the specific type of inputs and outputs'.

[231] For a discussion of the standard 'the most advanced models' see forthcoming commentary on Article 3(64) in this work.

[232] For article 51(1)(b)'s complementary purpose, see Section 2.1.2.

[233] See forthcoming commentary on Article 3(64) in this work.

[234] For classification under article 51(1)(b) based on a model's impact, see Section 2.1.2.1.4.

[235] See Section 2.1.1.2.

[236] See Section 2.1.2.2.

identical effect, in particular the application of Article 55(1) obligations, that follows from classification under both provisions.[237] Additionally, the reference in point (d) of Annex XIII to modality-specific high-impact capabilities signals that the threshold of high-impact capabilities retains relevance for Article 51(1)(b), even if only with regard to specific domains. The more compelling view is therefore that capabilities below the Article 51(1)(a) threshold are not sufficient for meeting Article 51(1)(b)'s standard of capabilities equivalent to high-impact capabilities in themselves. Rather, such lower capabilities would need to be offset by other factors, including domain-specific high-impact capabilities,[238] a particularly high impact[239] or the prospect that the model will predictably reach high-impact capabilities soon.

2.1.2.1.4. Equivalent impact

58. Article 51(1)(b) allows for classification based not only on a GPAI model's "equivalent capabilities"[240] but also based on its impact. The impact of a GPAI model can be described in terms of its reach, including the number of its users, and in terms of any actual or reasonably foreseeable negative effects that stem from it.[241] Points (f) and (g) of Annex XIII contain criteria particularly relevant for assessing model's impact.[242]

59. It has been questioned whether a model's impact alone can form the basis for a model's classification as presenting systemic risk under Article 51(1)(b),[243] as Article 3(65) defines systemic risk as 'specific to high-impact capabilities' and not as specific to a model's high impact.[244] In particular, it has been argued that a model's widespread use alone cannot be sufficient for classification.[245] While this is ultimately correct, it must be noted, however, that a model's reach is indeed a relevant factor to determine a model's risk, as widespread use statistically increases the overall probability of harm occurring.[246] Indeed, according to Recital 110 and the Safety and Security Chapter of the Code of Practice, the risk associated with a model increases *both* with its capabilities and with its reach.[247] At the same time, the reference to actual or reasonably foreseeable negative

[237] See Section 2.1.4.
[238] See Section 2.1.2.1.3.1.
[239] See Section 2.1.2.1.4.
[240] See Section 2.1.2.1.3.
[241] See AI Act, art 3(65) ('having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole'); see also AI Act, annex XIII, point (f) and (g). For an in-depth discussion of the concept of impact see commentary on Article 3(65) in this work.
[242] See Bond and Abbady, 'Art. 52' (n 70) 833–834, s 3.2. Other criteria listed in annex XIII, such as the number of model parameters (point (a)) or the quality and size of the data set (point (b)), appear to be less strong indicators of a model's impact and rather related to the model's capabilities. This raises the question of whether and to what extent the Commission must take all of annex XIII's criteria into account for determining whether a model has an impact equivalent to high-impact capabilities under article 51(1)(b) (see Section 2.1.2.1.2.1., para 41).
[243] See Bond and Abbady, 'Art. 51' (n 138) 833–834, s 3.2; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 32.
[244] Bond and Abbady, 'Art. 51' (n 138) 834, s 3.2.
[245] Bond and Abbady, 'Art. 51' (n 138) 833–834, s 3.2 ('Nonetheless, it does not appear likely that the legislators intended for the possibility of a GPAI model qualifying as a systemic risk model for the mere reason that it is widely used (and thus potentially has significant reach), however. The definition of "systemic risk" in the AI Act after all is: a risk that is *specific to the high-impact capabilities* of GPAI models [...].', emphasis by authors); Bernsteiner and Schmitt, 'Art. 51' (n 6) para 32.
[246] Bernsteiner and Schmitt, 'Art. 51' (n 6) para 32; see AI Act, art 3(2): '"risk" means the combination of the probability of an occurrence of harm and the severity of that harm'.
[247] AI Act, recital 110, second sentence; Code of Practice Safety and Security Chapter (n 22) app 1.2.2.

effects alongside reach as contributing factors that determine a model's impact in Article 3(65)[248] demonstrates that the legislature does not equate impact with reach. This suggests that a model lacking both high-impact capabilities and the ability to produce significant negative effects on the protected interests set out in Article 3(65) cannot be classified as a GPAI model with systemic risk under Article 51(1)(b), even if it is widely used and therefore has a particularly high reach.

60. Regarding the standard for impact-based classification under Article 51(1)(b), the AI Act does not expressly specify when a model's impact is 'equivalent to those set out in point (a)'. While Article 51(1)(a) sets out the standard of high-impact capabilities, a direct comparison of a model's *impact* with this *capabilities* standard is not readily feasible. Therefore, it appears that the legislature used 'an impact equivalent to those set out in point (a)' as shorthand for 'an impact of a GPAI model being equivalent to the impact of GPAI models with high-impact capabilities'.[249]

61. On this basis, Article 51(1)(b) requires the determination of an 'impact standard' of GPAI models with high-impact capabilities against which models to be classified under Article 51(1)(b) can be evaluated to determine their impact equivalence. On the basis of Article 3(64) and (65)'s definitions of high-impact capabilities and systemic risk,[250] this likely involves a comparison of the model with the most advanced GPAI models in terms of their reach and their actual and reasonably foreseeable negative effects. The question of whether Article 51(1)(b)'s use of 'equivalent' requires a model to at least match the impact of GPAI models with high-impact capabilities,[251] or whether – since it does not require an 'equal' impact – it sets a slightly lower threshold, likely has less practical relevance, given that such distinctions are difficult to draw for a qualitative threshold such as impact.

62. The business user threshold in point (f) of Annex XIII is likely to assume considerable importance for impact-based classification under Article 51(1)(b). According to this provision, a model is presumed to have a high impact on the internal market due to its reach 'when it has been made available to at least 10 000 registered business users established in the Union'.[252] A model that reaches this threshold and is therefore presumed to have a high impact generally satisfies the

[248] See AI Act, art 3(65): 'having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole'.

[249] Article 51(1)(b)'s wording may be due to the fact that the legislature sought to include both capabilities-based and impact-based classification in one classification condition.

[250] See AI Act, art 3(64) ('capabilities that match or exceed the capabilities recorded in the most advanced models') and AI Act, art 3(65) ('having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole')

[251] See AI Act, art 3(64).

[252] The AI Act does not define a 'registered business user'. For a discussion of its definition see Haar and Siglmüller, 'Art. 51' (n 9) paras 74–76 who argue that this criterion relates to the business and not employees of the business and rejects the possibility that machines or computer systems are users; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 26 fn 50 arguing a user needs to operate individually in the market in order to qualify as a business user. The DMA defines a 'business user' as 'any natural or legal person acting in a commercial or professional capacity using core platform services for the purpose of or in the course of providing goods or services to end user' (DMA art 2(21)), whereas the P2B Regulation defines a 'business user' as 'any private individual acting in a commercial or professional capacity who, or any legal person which, through online intermediation services offers goods or services to consumers for purposes relating to its trade, business, craft or profession' (Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services [2019] OJ L 186/57 ("P2B Regulation"), art 2(1)). For a discussion of these definitions, see Jan-Frederick Göhsl and Daniel Zimmer 'VO (EU) 2022/1925 Art. 2 Begriffsbestimmungen' in Torsten Körber, Heike Schweitzer and Daniel Zimmer (eds.), *Immenga/Mestmäcker Wettbewerbsrecht Band 1: EU Kommentar zum Europäischen Kartellrecht* (7th edn C H Beck 2025) para 78.

requirements to be classified as presenting systemic risk under Article 51(1)(b).[253] A high impact linguistically forms part of the notion of 'high-impact capabilities' which suggests – assuming that the legislature chose the term 'high-impact capabilities' deliberately[254] – that a model with a (presumed) high impact also has an impact equivalent to high-impact capabilities under Article 51(1)(b).[255] Recital 111 confirms this interpretation, stating that 'a general-purpose AI model should be considered to present systemic risks if it has [...] significant impact on the internal market due to its reach'.[256] The fact that this threshold, unlike the training compute threshold under Article 51(2), is not contained in Article 51 itself[257] and does not precisely correspond to Article 51(1)(b)'s wording of 'an impact equivalent to those set out in point (a)' does not undermine this conclusion, as the criteria in Annex XIII, and thus the high-impact presumption under point (f) of Annex XIII, evidently relate to Article 51(1)(b),[258] and the presumption would otherwise appear to lack a meaningful scope of application.[259]

63. The AI Act does not specify whether the presumption of a model's high impact under point (f) of Annex XIII is rebuttable. Its refutability is supported by the fact that a high number of registered business users does not necessarily indicate an impact sufficient to assume the presence of systemic

[253] See Wachter (n 2) 715, fn 202 ('Annex XIII(f) assumes systemic risk when the model is made available to 10,000 registered business users.'); Schneider and Schneider, 'Art. 51' (n 4) para 25; opposing view: Bernsteiner and Schmitt, 'Art. 51' (n 6) para 32 arguing that a high reach does not imply a systemic risk *per se*; Bond and Abbady, 'Art. 51' (n 138) 834, s 3.2, assuming that 'mere widespread use cannot be a deciding factor for determining whether a GPAI model entails systemic risk'.

[254] There is only limited information available about the origin of this term which has been introduced during the trilogue (see Jonathan Kirschke-Biller and Anna Lena Füllsack, 'Art. 3 Begriffsbestimmungen' in Jens Schefzig and Robert Kilian (eds), *Beck'scher Online-Kommentar KI-Recht* (4th edn, C H Beck 2025) para 728). Given that the term has no discernible precedent and is rather specific (a more generic term such as 'frontier AI capabilities' would have been conceivable as well), this indicates that the legislature indeed associated high-impact capabilities with high impact.

[255] It is a well-established principle of interpretation that the wording of article or section titles influences the interpretation of operative EU law provisions (*Schrems II* (n 10) para 92; *Papasavvas* (n 10) para 39). It appears possible to extend this principle, by analogy, to the denomination chosen for a technical term defined by EU legislation such as 'high-impact capabilities'.

[256] See Bond and Abbady, 'Art. 51' (n 138) 834, s 3.2, who question whether impact is a criterion for classification independent of capabilities but concede that this 'is also supported by Recitals 110 and 111, which respectively provide that systemic risks posed by GPAI models should be understood to increase with model capabilities and model reach, and that that a GPAI model should be understood to present systemic risk if it has significant impact on the market due to its reach.'; see also AI Act, recital 110 which states '[s]ystemic risks should be understood to increase with [...] model reach'. The different terminology ('high impact' in point (f) of annex XIII and 'significant impact' in article 3(65) and recital 111) appears to be a drafting inconsistency rather than a significant distinction as it is not apparent how these two would be differentiated.

[257] An annex forms an integral part of a EU legislative act (see Case 222/81 *BAZ Bausystem AG v Finanzamt München für Körperschaften* [1982] ECLI:EU:C:1982:256 para 7).

[258] The placement of the business user threshold in point (f) of annex XIII can be attributed to the legislative choice to provide for criteria relevant for article 51(1)(b)'s condition not in the provision itself but in an annex; see also Section 2.1.2.1.2.

[259] Principles of systematic interpretation therefore argue in favour of a model's high impact being sufficient for classification under Article 51(1)(b), as one can assume that the legislature would not have included a presumption largely devoid of legal effect (see, for example, *RW v Österreichische Post AG* (n 159) para 29 and the case-law cited therein ('[W]here a provision of EU law is open to several interpretations, preference must be given to that interpretation which ensures that the provision retains its effectiveness.'); for a discussion of this aspect of systematic interpretation see Lenaerts and Gutiérrez-Fons (n 211) 17–21).

risks, since reach is, as laid out above,[260] only one relevant factor to determine a model's impact.[261] Moreover, an irrefutable presumption would conflict with the Commission's duty to have regard to Annex XIII for designation under Article 51(1)(b) in conjunction with the first subparagraph of Article 52(4).[262] This duty likely implies that the Commission must consider all of the Annex XIII criteria even if it does not decisively base its decision on all of them,[263] thereby precluding any single criterion – as would be the case with an irrefutable presumption – from automatically qualifying the model for designation.[264]

64. The business user threshold in point (f) of Annex XIII is not the only criterion the Commission has to take into account for impact-based classification under Article 51(1)(b). Point (g) of Annex XIII mentions an additional reach-related criterion – the model's number of registered end users – but refrained from combining this criterion with a presumption or any other threshold indicating a relevant number of registered end users.[265] Interestingly, Annex XIII does not contain any criterion specifically aimed at determining the actual or reasonably foreseeable negative effects stemming from a model. These effects may require consideration under point (g) of Annex XIII mentioning 'evaluations of capabilities of the model'.[266] More generally, it appears unlikely that Annex XIII restricts the Commission in its consideration of a model's effects, as Article 3(65) expressly mentions these in its reference to a model's impact[267] and Annex XIII's list of criteria is arguably non-exhaustive.[268] The assessment of a model's effects could, for example, be informed by so-called AI incident trackers and databases.[269]

---

[260] See para 59.

[261] See Schneider and Schneider, 'Art. 51' (n 4) para 16 arguing that no single criterion contained in annex XIII necessarily qualifies a model for designation; see also Bernsteiner and Schmitt, 'Art. 51' (n 6) para 32; Bond and Abbady, 'Art. 51' (n 138) 834, s 3.2.

[262] For this duty, see Section 2.1.2.1.2.1.

[263] See Section 2.1.2.1.2.1.

[264] See Schneider and Schneider, 'Art. 51' (n 4) para 16.

[265] Carey (n 182) 9 mentions the number of one million registered end users without further explanation ('[H]igh impact may be determined through the number of registered business users (10000) or number of registered end-users (1000000) in the internal market.'). To determine which number of registered end users could be indicative of a model's high impact, regard could be had to statistics of the number of registered business and end users in GPAI models with systemic risk in general, insofar as such information is available. Article 3(2)(b) of the DMA mentions the number of 45 million monthly active end users established or located in the Union alongside the number of 10,000 yearly active business users established in the Union. These figures provide a first indication of a normal ratio of end users to business users that the legislature may assume in certain digital markets. However, the DMA's number of 45 million monthly active end users cannot be adopted as a number of registered end users indicative of a model's high impact under the AI Act in light of the diverging regulatory contexts.

[266] While a model's effects are conceptually different from its capabilities, their evaluation will often overlap in practice as a model's capabilities both determine its possible effects and are often better understood through real-world incidents. The connection is particularly evident for capabilities which are defined in terms of the potential to cause a certain effect, such as the 'capabilities that could cause the persistent and serious infringement of fundamental rights' or 'capabilities to manipulate, persuade or deceive' mentioned in the Code of Practice (see Code of Practice Safety and Security Chapter (n 22) app 1.3.1, points (3) and (4)).

[267] See AI Act, art 3(65). See also AI Act, recital 97, thirteenth sentence: 'Considering their potential significantly negative effects, the general-purpose AI models with systemic risk should always be subject to the relevant obligations under this Regulation.'

[268] See Section 2.1.2.1.2.2.

[269] See, for example, 'AI Incident Database' <https://incidentdatabase.ai/> accessed 7 January 2026 and MIT's 'AI Incident Tracker' (*MIT AI Risk Initiative*) <https://airisk.mit.edu/ai-incident-tracker> accessed 7 January 2026.

2.1.2.1.5. Cumulative equivalence

65. The preceding discussion has examined the conditions under which a model's capabilities or impact might independently satisfy Article 51(1)(b)'s standard of equivalence to high-impact capabilities. The question thus arises of whether capabilities and impact, when individually insufficient, could together constitute a basis for equivalence and thereby support classification under Article 51(1)(b) ("cumulative equivalence").[270]

66. Article 51(1)(b)'s wording is ambiguous in this respect. The phrase 'capabilities or an impact equivalent to those set out in point (a)' may be read as referring to *either* capabilities that are equivalent to high-impact capabilities *or* an impact that is equivalent to high-impact capabilities. At the same time, the provision's wording does not exclude a reading under which 'equivalent to those set out in point (a)' attaches to 'capabilities or an impact' as a composite expression.[271] On one reading, therefore, at least one element – either capabilities or impact – must independently establish equivalence to high-impact capabilities, such that their mere combination cannot suffice where each falls short on its own ("independent equivalence"). On an alternative reading, capabilities and impact may be considered together, with their cumulative effect potentially establishing the requisite equivalence even where neither alone would meet the threshold.[272]

67. A reading of Article 51(1)(b)'s wording as requiring "independent equivalence" appears more natural. However, the fact that Annex XIII does not expressly distinguish between (i) criteria relevant to assessing whether a GPAI model has sufficient capabilities and (ii) those relevant to assessing whether it has sufficient impact rather supports an interpretation of Article 51(1)(b) of allowing classification in cases of cumulative equivalence. Article 51(1)(b)'s general reference to Annex XIII, combined with the existence of a single undivided list in that Annex, could suggest that in principle all criteria may be considered together in determining whether a model meets Article 51(1)(b)'s requirements. The force of this argument is, however, weakened by the observation that it appears possible to identify a division whereby the criteria in points (a) to (e) of Annex XIII primarily relate rather to a model's capabilities, while the criteria in points (f) and (g) primarily relate to a model's impact.[273]

68. Moreover, the relevance of cumulative equivalence under Article 51(1)(b) finds some (albeit limited) support in Recital 111's twelfth sentence, which states that designation decisions under Article 51(1)(b) should be taken 'on the basis of an overall assessment' of the Annex XIII criteria.

---

[270] This question has so far received limited attention in legal scholarship. Notably, however, many authors emphasise that a Commission decision in the context of article 51(1)(b) requires an *overall assessment* of the criteria contained in annex XIII (see Hofmann-Coombe (n 6) para 49; Schneider and Schneider, 'Art. 51' (n 4) para 15; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 27; Haar and Siglmüller, 'Art. 51' (n 9) para 60). This emphasis suggests that these authors do not perceive the question of whether the model's equivalence is based on its capabilities or its impact as relevant.

[271] Such a reading would only be excluded by the use of 'either ... or ...' ('either capabilities or an impact equivalent to those set out in point (a)') or the repetition of the equivalence clause ('capabilities equivalent to those set out in point (a) or an impact equivalent to those set out in point (a)').

[272] If 'equivalent to those set out in point (a)' attaches to 'capabilities or an impact' as a composite expression, this suggests that the use of 'or' in article 51(1)(b) permits aggregation. For the ambiguity surrounding the use of 'or' in legal drafting see European Parliament, Council of the European Union and European Commission, 'Joint Handbook for the Presentation and Drafting of Acts Subject to the Ordinary Legislative Procedure' (2023) <https://www.consilium.europa.eu/media/67390/joint_handbook_en_01-october-2023_clean_def_final.pdf> s D.4.4.1 ('The conjunction "or" should be used alone only when the nature of the link is clear because, as the Court has held, the meaning of this conjunction differs depending on the context in which it is used.' as well as Case C-304/02 *Commission of the European Communities v French Republic* [2005] ECLI:EU:C:2005:444 para 83.

[273] See Section 2.1.2.1.3. and Section 2.1.2.1.4.

In light of this recital, one could argue that Article 51(1)(b)'s main substantive requirements are not capabilities or impact as such but rather equivalence to high-impact capabilities. However, as laid out above,[274] there are grounds for not giving the recital's reference to an 'overall assessment' too much weight as recitals may clarify the legislature's intention but do not have binding legal force[275] and a requirement of an overall assessment does not appear in the text of Article 51(1)(b) itself.[276]

2.1.2.1.6. Interactions with Article 51(1)(a)

69. Having established that Article 51(1)(b) sets out substantive requirements that are distinct from Article 51(1)(a) and having analysed these distinct requirements, two related questions arise concerning the interaction between Article 51(1)(a) and (b): first, is it sufficient for classification under Article 51(1)(b) that the model has high-impact capabilities?[277] Second, can a provider prevent designation under Article 51(1)(b) by proving its model lacks high-impact capabilities?[278] Both questions turn on how Article 51(1)(b) operates in cases where the presence or absence of high-impact capabilities has been definitively established.

2.1.2.1.6.1. Presence of high-impact capabilities

70. Article 51(1)(b)'s wording suggests that a GPAI model with high-impact capabilities meets the substantive requirements for classification under Article 51(1)(b), as high-impact capabilities are by definition equivalent to themselves.[279] A systematic argument can, however, be made against designation of a GPAI model with high-impact capabilities as presenting systemic risk on the basis of Article 51(1)(b) in conjunction with Article 52(4)'s first subparagraph. Such models are automatically classified as presenting systemic risk under Article 51(1)(a).[280] Moreover, an instance where designation of such automatically classified models may be valuable – namely, to make a legally binding determination that a model meets Article 51(1)(a)'s classification condition in the absence of provider notification pursuant to Article 52(1)'s first sentence – is arguably already addressed by Article 52(1)'s third sentence establishing the Commission's power to designate GPAI models with high-impact capabilities as presenting systemic risk.[281] In light of this provision, it would appear redundant to allow the Commission to designate a GPAI model with high-impact capabilities as presenting systemic risk on the basis of Article 51(1)(b) in conjunction with Article 52(4)'s first subparagraph. The legislative choice to limit the scope of Article 52(5)'s procedure to contest designation to designation pursuant to Article 52(4)'s first subparagraph additionally suggests that the Commission may not use its powers under Article 52's designation provisions interchangeably.[282]

2.1.2.1.6.2. Absence of high-impact capabilities

71. Article 3(65) characterises systemic risk as 'specific to the high-impact capabilities of general-purpose AI models'. This raises the question of whether a provider can prevent its model's

---

[274] See Section 2.1.2.1.2.1.
[275] See *Puppinck* (n 12) paras 75–76.
[276] See also Section 2.1.2.1.2.1.
[277] See Section 2.1.2.1.6.1.
[278] See Section 2.1.2.1.6.2.
[279] To the same effect: Bernsteiner and Schmitt, 'Art. 51' (n 6) para 24–25.
[280] See Section 2.1.1.1.
[281] See commentary on Article 52, Section 2.1.3. in this work.
[282] See AI Act, art 52(5), first sentence: 'Upon a reasoned request of a provider whose model has been designated as a general-purpose AI model with systemic risk pursuant to paragraph 4, [...]'. For a discussion of article 52(5)'s scope, see commentary on Article 52, Section 2.4.1. in this work.

classification under Article 51(1)(b) by proving that it does not have high-impact capabilities. Under a literal reading, 'specific to' can either mean *exclusive to* – implying that only GPAI models with high-impact capabilities can present systemic risks – or *characteristic of* – implying that GPAI models with high-impact capabilities typically present systemic risks without excluding that GPAI models without such capabilities may under certain circumstances present systemic risks as well. Arguments for and against each interpretation are discussed elsewhere.[283] For the present discussion, it is interesting to consider the implications of different interpretations of Article 3(65)'s definition of systemic risk.

72. Where 'specific to' is read as *exclusive to*, the classification of a GPAI model without high-impact capabilities under Article 51(1)(b) appears hard to justify, as the model would – by definition – not present systemic risks. Where it is read as *characteristic of*, however, the possibility to classify a GPAI model without high-impact capabilities under Article 51(1)(b) could not only be justified but even warranted. In that case, if Article 51(1)(b) would not allow for classification of GPAI models without high-impact capabilities that nonetheless present systemic risks, these models would escape classification, given that Article 51(1)(a) permits only high-impact capabilities-based classification and the existence of further classification pathways outside of Article 51(1) appears at least very uncertain.[284] It appears doubtful whether the existence of GPAI models presenting systemic risk that cannot be classified as such can be reconciled with the AI Act's rules for GPAI models with systemic risk, as that would mean that these systemic risks could present themselves on the Union market without thorough assessment or mitigation.[285]

2.1.2.2. Commission designation in the context of Article 51(1)(b)

73. Under Article 51(1)(b) – in contrast to Article 51(1)(a)[286] – models are not classified as presenting systemic risk automatically. Rather, classification under Article 51(1)(b) is 'based on a decision of the Commission' according to which the GPAI model has capabilities or an impact equivalent to those set out in Article 51(1)(a), that is, high-impact capabilities. According to Recital 111's twelfth sentence, this decision should be taken 'on the basis of an *overall assessment*' (emphasis added) of the criteria listed in Annex XIII,[287] a requirement which does not appear in the text of Article 51(1)(b) itself.[288] Absent a Commission decision, models that meet Article 51(1)(b)'s substantive requirements are not considered GPAI models with systemic risk.

74. Although Article 51(1)(b) contains no express reference to Article 52(4)'s first subparagraph, compelling reasons suggest that the Commission decision referred to under Article 51(1)(b) constitutes a designation decision in the sense of Article 52(4)'s first subparagraph.[289] Both

[283] See forthcoming commentary on Article 3(65) in this work; for a discussion of 'specific to' under article 3(65), see also Hacker, Kasirzadeh and Edwards (n 61) 24–25.

[284] See Section 2.1.3.

[285] See AI Act, art 55(1)(a); see also AI Act, recital 97, thirteenth sentence: 'Considering their potential significantly negative effects, the general-purpose AI models with systemic risk should always be subject to the relevant obligations under this Regulation.'

[286] See Section 2.1.1.1.

[287] See Hofmann-Coombe (n 6) para 49; Schneider and Schneider, 'Art. 51' (n 4) para 15; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 27; Haar and Siglmüller, 'Art. 51' (n 9) para 60; Lukas Feiler, Nikolaus Forgó and Michaela Nebel, 'Article 51' in *The EU AI Act: A Commentary* (Globe Law and Business 2025) para 8.

[288] Recitals may clarify the legislature's intention but do not have binding legal force, see *Puppinck* (n 12) paras 75–76.

[289] Haar and Siglmüller, 'Art. 51' (n 9) para 59; Haar and Siglmüller, 'Art. 52' (n 24) para 19; Feiler, Forgó and Nebel, 'Article 51' (n 287) para 9; Hecht (n 39) 34. The Commission Guidelines (n 39) para 42 classify the decision

provisions' shared reference to Annex XIII as well as Recital 111's characterisation of the Article 51(1)(b) decision as a designation decision[290] strongly support this reading.[291]

75. The procedural rights laid down in Article 18 MSR apply *mutatis mutandis* for a provider facing designation of its model under Article 52(4)'s first subparagraph,[292] at least from 2 August 2026.[293] In particular, the provider has the right to be heard under Article 18(3) MSR.[294]

2.1.3. Further classification pathways

76. Intuitively, and by the article's title 'Classification rules',[295] Article 51(1) seems to exhaustively list all conditions under which a GPAI model may be classified as presenting systemic risk. This section examines whether the AI Act, and particularly Article 52(1)'s third sentence, captures instances in which systemic risks materialise in a GPAI model with neither high-impact capabilities nor equivalent capabilities or impact, thereby establishing alternative pathways to classification independent of those provided by Article 51(1). Convincing arguments speak against the existence of such alternative classification pathways beyond Article 51(1).[296] Yet, in the subsequent Article 52 on 'Procedure', there are two provisions that establish the Commission's power to designate GPAI models as presenting systemic risk without directly referring to Article 51(1)(a) or (b) and their respective requirements.[297] As such, neither Article 52(1)'s third sentence nor Article 52(4)'s first subparagraph *expressly* require that a GPAI model has 'high-impact capabilities' under Article 51(1)(a) or 'capabilities or an impact equivalent to [high-impact capabilities]' under Article 51(1)(b) for its designation as presenting systemic risk.

77. One can make sense of this absence of an express reference to Article 51(1)'s requirements in broadly two ways. First, this omission could be unintentional, meaning that these designation provisions do implicitly rely on the requirements for classification under Article 51(1)(a) and (b).[298] Alternatively, however, the lack of a reference to Article 51(1) could have been deliberate – and thus of significance – implying that these designation provisions would establish independent

---

under article 51(1)(b) as a designation without expressly linking it to article 52(4)'s first subparagraph; opposing view: Hofmann-Coombe (n 6) paras 48, 51 arguing that the decision that the model has capabilities or an impact equivalent to those set out in point (a) and the decision that classifies the model as presenting systemic risk are separate decisions in a two-stage procedure; see also commentary on Article 52, Section 2.1.3.1. in this work.

[290] AI Act, recital 111, eleventh sentence ('To complement this system, there should be a possibility for the Commission to take individual decisions designating a general-purpose AI model as a general-purpose AI model with systemic risk if it is found that such model has capabilities or an impact equivalent to those captured by the set threshold.').

[291] For an in-depth analysis of the arguments in favour of the Commission decision referred to under article 51(1)(b) constituting a designation decision in the sense of article 52(4)'s first subparagraph, see commentary on Article 52, Section 2.1.3.1. in this work.

[292] Hilgendorf and Härtlein, 'Art. 52' (n 11) para 10; see commentary on Article 52, Section 2.1.3.2. in this work.

[293] According to article 113(2) and (3)(b), article 52 applies from 2 August 2025, whereas article 94 only applies from 2 August 2026. During this transitional period, a provider's procedural rights, including the right to be heard, may derived from article 41(2)(a) of the Charter.

[294] Hilgendorf and Härtlein, 'Art. 52' (n 11) para 10; to the same effect on the basis of article 41(2)(a) of the Charter: Bernsteiner and Schmitt, 'Art. 52' (n 17) para 21; see also Section 2.1.3.2.

[295] For the role of the wording of an article's title in the interpretation of operative provisions, see *Schrems II* (n 10) para 92; see also *Papasavvas* (n 10) para 39 with regard to a section title.

[296] See commentary on Article 52, Section 2.1.3.3.2. in this work.

[297] AI Act, art 52(1), third sentence, and art 52(4), first subparagraph; for these provisions see commentary on Article 52, Section 2.1.3. and 2.3.1. in this work respectively.

[298] See Haar and Siglmüller, 'Art. 51' (n 9) paras 18, 21.

pathways for designation of a GPAI model as presenting systemic risk. Of course, more nuanced variations of these two broad interpretations are conceivable as well.[299]

78. Under the first interpretation – that of Article 52 implicitly referring back to Article 51 – Article 52(1)'s third sentence would relate to Article 51(1)(a)'s high-impact capabilities-based classification, while Article 52(4)'s first subparagraph would correspond to classification under Article 51(1)(b), which is based on capabilities or impact equivalent to high-impact capabilities.[300] The arguments that support this reading are explored more extensively elsewhere.[301] In short, the argument mainly relies on the positioning of Article 52(1)'s third sentence immediately after the provisions on the notification obligation clearly relating to Article 51(1)(a), and the broad congruence of the wording of Article 52(4)'s first subparagraph and Article 51(1)(b).[302] In line with this interpretation, it can be argued that designation under Article 52(1)'s third sentence requires the GPAI model to have high-impact capabilities[303] and that designation under Article 52(4)'s first subparagraph requires the GPAI model to have capabilities or an impact equivalent to high-impact capabilities.[304] This interpretation is indeed convincing and shared by the Commission Guidelines.[305]

79. However, the alternative interpretation – that these designation provisions establish independent classification pathways – merits consideration, particularly regarding Article 52(1)'s third sentence.[306] This provision requires the Commission to become aware of 'a general-purpose AI model presenting systemic risk' without reference to high-impact capabilities.[307] Arguments centring on this wording support reading this provision as requiring the model to present risks that fall under the definition of systemic risk under Article 3(65).[308]

80. Such additional classification pathways could play a role in the classification framework's adaptability to evolving technological developments.[309] Article 51(1)'s framework relies heavily on high-impact capabilities as a proxy for systemic risk. This particularly applies to classification under Article 51(1)(a), which directly requires such high-impact capabilities. It applies to a lesser extent to classification under Article 51(1)(b) as well, which requires capabilities or an impact equivalent to

[299] These would include interpretations such as only one of the designation provisions under article 52 relating to article 51(1), or the designation provisions under article 52 allowing for designation in cases where one of conditions under article 51(1) is met but also under additional conditions.

[300] See Haar and Siglmüller, 'Art. 51' (n 9) para 18; see commentary on Article 52, Section 2.1.3.1. in this work.

[301] See commentary on Article 52, Section 2.1.3.1. in this work.

[302] See commentary on Article 52, Section 2.1.3.1. in this work.

[303] Commentary on Article 52, Section 2.1.3.3. in this work.

[304] Commentary on Article 52, Section 2.3.1. in this work.

[305] Commission Guidelines (n 39) para 43; see also: Haar and Siglmüller, 'Art. 51' (n 9) paras 18, 21; Haar and Siglmüller, 'Art. 52' (n 24) para 19.

[306] The arguments for designation under article 52(4)'s first subparagraph relating to article 51(1)(b) and not establishing an independent classification pathway appear particularly convincing (see commentary on Article 52, Section 2.3.1. in this work). For this reason, the remainder of this paragraph focuses on designation under article 52(1)'s third sentence.

[307] AI Act, art 52(1), third sentence.

[308] See commentary on Article 52, Section 2.1.3.3. in this work.

[309] Concerns regarding future-proofness and resilience to disruption in light of technological developments constitute a recurrent theme in the AI Act's recitals and have motivated various of its provisions (see, for example, AI Act, recital 12, first sentence; recital 101, last sentence; recital 138, second sentence; recital 179, seventh sentence). As laid out Section 2.1.2.1. para 33, it appears plausible that the legislature, by creating a more flexible Article 51(1)(b) alongside Article 51(1)(a), sought to ensure that the classification framework provided by both provisions would be future-proof and resilient to disruption.

high-impact capabilities.[310] This approach fits with how the AI Act links the notion of systemic risk on a *definitional level* to the concept of high-impact capabilities, defining it as 'specific to the high-impact capabilities of general-purpose AI models'.[311] However, it may be less appropriate on the *level of classification*. There is considerable uncertainty regarding future technological developments and the ability of evaluation methods to keep up with the pace of capability improvements. Presently, the determination of high-impact capabilities relies heavily on the training compute threshold in Article 51(2),[312] but training compute could become a less meaningful proxy for capabilities as AI development methods evolve.[313] Where one cannot reliably evaluate whether a model has high-impact capabilities, having classification pathways that function independently from capability assessments would be crucial. This could particularly be the case in scenarios where clear evidence that systemic risks have already materialised is available – for example, because of the occurrence of serious incidents – but strong approximations for a model's capabilities are not available.[314] Besides impact-based classification under Article 51(1)(b),[315] an additional systemic risk-based classification pathway under Article 52(1)'s third sentence could fulfil this function.

81. Strong arguments exist against such additional classification pathways, however. Although Article 51(1)'s wording does not expressly exclude the existence of further classification pathways,[316] it provides the main argument against them. If the drafters had intended to create pathways beyond Article 51(1)(a) and (b), the logical place for such provisions would have been within the classification provision itself. Furthermore, impact-based classification under Article 51(1)(b) may already permit classification based on actual or reasonably foreseeable negative effects of a model,[317] which would render an additional classification pathway largely redundant.

2.1.4. Effects of classification

82. A model's classification under Section 1 of Chapter V has the effect that the AI Act's provisions for GPAI models with systemic risk apply.[318] The legislature presumably considered this so apparent

[310] For the relevance of the concept of high-impact capabilities for classification under article 51(1)(b) see, in particular, Section 2.1.2.1.3. and Section 2.1.2.1.4. Nonetheless, convincing arguments support an interpretation of article 51(1)(b) establishing substantive requirements for classification distinct from those under article 51(1)(a), a view which is, however, contested in legal scholarship (see Section 2.1.2.1.1.).

[311] See AI Act, art 3(65).

[312] See Commission Guidelines (n 39) paras 29, 31; Haar and Siglmüller, 'Art. 52' (n 24) para 6. This may change over time as article 51(1)(a) and (3) envisage the adoption of indicators and benchmarks that help evaluate the model's high-impact capabilities via delegated act by the Commission.

[313] Recital 111's sixth sentence acknowledges this by stating that '[a]ccording to the state of the art *at the time of entry into force of this Regulation*, the cumulative amount of computation used for the training of the general-purpose AI model measured in floating-point operations is one of the relevant approximations for model capabilities' (emphasis added). See Venkat Somala, Anson Ho and Séb Krier, 'Three Challenges Facing Compute-Based AI Policies' (2025) <https://epoch.ai/gradient-updates/three-issues-undermining-compute-based-ai-policies> accessed 7 January 2026.

[314] See also AI Act, recital 97, thirteenth sentence: 'Considering their potential significantly negative effects, the general-purpose AI models with systemic risk should always be subject to the relevant obligations under this Regulation.'

[315] See Section 2.1.2.1.4.

[316] Article 51(1) does not provide that a model 'shall *only* be classified' as presenting systemic risk if it meets any of the conditions under article 51(1)(a) or (b).

[317] See Section 2.1.2.1.4.

[318] Specifically with respect to the obligations that follow from classification: Commission Guidelines (n 39) para 44; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 3; however, see Hacker and Holweg (n 21) 5.

that it refrained from specifically providing for it.[319] Nevertheless, this consequence follows from a number of textual indicators, including the positioning of the classification rules at the beginning of Chapter V of the AI Act, the correspondence in wording between Article 51(1) and Article 55(1),[320] and Article 52(5) and (3) implying that classification leads to the GPAI model being 'considered to present systemic risks'.

83. Most notably, providers of GPAI models with systemic risk face *obligations through Article 55(1)*, requiring them to perform model evaluations, assess and mitigate systemic risks associated with the model, report serious incidents and ensure an adequate level of cybersecurity protection.[321] Beyond these obligations, special provisions for GPAI models with systemic risk and their providers appear throughout Section 2 ('Obligations for Providers of General-Purpose AI Models') of Chapter V, comprising *Articles 53 and 54*, together with the related Annex XI. Under Article 53(1)(a) in conjunction with the second section of Annex XI, providers of GPAI models with systemic risk must draw up, keep up-to-date and provide, upon request, to the AI Office and the national competent authorities, additional information relating to the technical documentation of the model.[322] Furthermore, under Article 53(2)'s second sentence and Article 54(6), providers of GPAI models with systemic risks do not benefit from exceptions for certain open-source GPAI models from relevant obligations under Articles 53 and 54.[323] Article 52(6) additionally provides for the publication of a list of GPAI models with systemic risk.[324]

84. Provisions specifically concerning GPAI models with systemic risk extend *beyond Chapter V* of the AI Act to Section 5 ('Supervision, Investigation, Enforcement and Monitoring in Respect of Providers of General-Purpose AI Models') of Chapter IX and Chapter XII ('Penalties') of the AI Act. Under Article 92(1)(b), the AI Office may conduct evaluations of GPAI models with systemic risk to investigate systemic risks at Union level;[325] under Article 93(1)(b), the Commission may request a provider to implement mitigation measures where an evaluation carried out in accordance with Article 92 has given rise to serious and substantiated concern of a systemic risk at Union level;[326] and under Article 101(1)(d), a provider of a GPAI model with systemic risk may be fined where it failed to make available to the Commission access to the model with a view to conducting such evaluations.[327] Beyond these provisions specific to GPAI models with systemic risk, the provisions for all GPAI models continue to apply to GPAI models with systemic risk.[328]

85. One may note that for the purpose of the abovementioned provisions, the classification rules under Section 1 of Chapter V establish the term *general-purpose AI model with systemic risk* and its

[319] The provisions for classification of GPAI models with systemic risk under the AI Act differ in this respect from the provisions for gatekeeper designation under the Digital Markets Act (see DMA, art 3(10)) and the provisions for the designation of very large online platforms and very large online search engines under the Digital Services Act (see DSA, art 33(1) and (6)).

[320] See AI Act, art 51(1) ('A general-purpose AI model shall be *classified as a general-purpose AI model with systemic risk* if [...]', emphasis added) and AI Act, art 55(1) ('In addition to the obligations listed in Articles 53 and 54, providers of *general-purpose AI models with systemic risk* shall: [...]', emphasis added).

[321] AI Act, art 55(1). These obligations are analysed in-depth in commentary on Article 55 in this work.

[322] See commentary on Article 53, paras 36–41 in this work.

[323] See commentary on Article 53, para 110 in this work and commentary on Article 54, para 48 in this work.

[324] See commentary on Article 52, Section 2.5. in this work.

[325] See forthcoming commentary on Article 92 in this work.

[326] See commentary on Article 93 in this work.

[327] See forthcoming commentary on Article 101 in this work.

[328] See, notably, AI Act, art 53.

variants[329] as a technical term whose meaning derives entirely from the classification process itself.[330] Where substantive provisions apply to GPAI models with systemic risk, the applicability of those provisions does not depend on an independent test of whether the GPAI model actually presents risks that fall under the systemic risk definition of Article 3(65).[331] Any challenge on these grounds must in principle target the model's classification itself through the procedure for contesting classification under Article 52(2) and (3), the procedure for contesting classification under Article 52(5) or an action for annulment before the Court of Justice of the European Union under Article 263(4) TFEU.[332] To permit otherwise would circumvent the specific requirements of these procedures and render systemic risk classification of any model perpetually open to challenge, which – as these very procedures demonstrate[333] – contradicts the purpose of classification of GPAI models as presenting systemic risk. Conversely, the Commission cannot enforce substantive obligations specific to GPAI models with systemic risk without a model's classification, requiring it to designate a GPAI model as presenting systemic risk where it is not automatically classified under Article 51(1)(a).

86. Unlike the provisions for gatekeeper designation under the DMA and for designation of very large online platforms (VLOPs) and very large online search engines (VLOSEs) under the DSA,[334] the AI Act does not provide for a transitional period between classification – either via automatic classification or through a designation decision – and the applicability of the obligations for GPAI models with systemic risk. A provider facing the prospect of classification of its model will therefore likely need to prepare to ensure compliance upon designation or automatic classification. Recital 112 clarifies in this context that a provider should be able to foresee its model's training compute threshold-based classification,[335] which would allow it to prepare accordingly.

---

[329] The AI Act at times uses 'general-purpose AI models *with systemic risks*' (see article 53(2), emphasis added) and 'general-purpose AI models [that] present systemic risks' (see article 54(6)) instead of 'general-purpose AI model *with systemic risk*' (see article 55(1), emphasis added). These inconsistencies are only partially reflected in different language versions of the AI Act (see article 53(2) and 55(1) of the French language version: 'modèles d'IA à usage général *présentant un risque systémique*') and appear to be unintentional.

[330] See Hofmann-Coombe (n 6) para 37; apparently opposing view: Hacker and Holweg (n 21) 5 who argue that GPAI models with systemic risk 'are defined by the AI Act, specifically Articles 3(64) and 3(65)' (without discussing the role of systemic risk classification under Section 1 of Chapter V of the AI Act in this context).

[331] See Hofmann-Coombe (n 6) para 37 who argues that article 51(1) prevails over article 3(65) ('[Die Vorschrift des Art. 51 Abs. 1 KI-VO] wirkt [...] wie eine Legaldefinition und verdrängt inhaltlich Art. 3 Nr. 65 KI-VO'); apparently opposing view: Hacker and Holweg (n 21) 5; further see para 82.

[332] For the procedures under article 52(2) and (3) and article 52(5), see commentary on Article 52, Section 2.2. and Section 2.4. in this work respectively. Commission decisions following contestation of classification as well as designation decision under article 52(1)'s third sentence and article 52(4)'s first subparagraph can be challenged under article 263(4) TFEU (see Bond and Abbady, 'Art. 52' (n 70) 842, s 3.2, 843, s 3.3.2 and 844, s 3.3.3; Haar and Siglmüller, 'Art. 52' (n 24) paras 26 and 27; Schneider and Schneider, 'Art. 52' (n 9) para 11); see also commentary on Article 52, Sections 2.1.3.5., 2.2.3.3., 2.3.1.2. and 2.4.3. in this work for Commission decisions under art 52(1), third sentence; (3); (4), first subparagraph; and (5) respectively).

[333] See, in particular, article 52(5)'s provisions on the timing of reassessment requests (see commentary on Article 52, Section 2.4.2.1. in this work).

[334] See DMA, art 3(10) ('The gatekeeper shall comply with the obligations laid down in Articles 5, 6 and 7 within 6 months after a core platform service has been listed in the designation decision pursuant to paragraph 9 of this Article.') and DSA, art 33(6), second subparagraph, second sentence ('The obligations set out in this Section shall apply, or cease to apply, to the very large online platforms and very large online search engines concerned from four months after the notification to the provider concerned referred to in the first subparagraph.')

[335] See AI Act, recital 112, fourth sentence: '[T]raining of general-purpose AI models takes considerable planning which includes the upfront allocation of compute resources and, therefore, providers of general-purpose AI models are able to know if their model would meet the threshold before the training is completed.' For automatic classification based on article 51(2)'s training compute threshold, see Section 2.2.1.

2.2. Article 51(2): Presumption of high-impact capabilities

87. Article 51(2) establishes a presumption of a GPAI model's high-impact capabilities based on the cumulative amount of computation used for its training ("training compute"), as it is '[a]ccording to the state of the art at the time of entry into force of this Regulation [...] one of the relevant approximations for model capabilities'.[336]

88. Under the enacted version of the AI Act, the presumption is triggered when training compute exceeds 10^{25} FLOPs,[337] a threshold that some authors claim 'is not based on empirical evidence but rather the result of a political compromise'.[338] According to a June 2025 estimate, 'over 30 publicly announced AI models from different AI developers' surpassed this threshold.[339] Given that training compute for frontier AI models has been estimated to have grown by more than four times per year between 2018 and 2024,[340] updates to Article 51(2) appear reasonable to expect,[341] and Article 51(3) empowers the Commission to update the training compute threshold accordingly.[342]

89. A GPAI model that meets this threshold is automatically classified as a GPAI model with systemic risk under Article 51(1)(a).[343] Article 51(2) draws no distinction based on when during the development process the threshold is met, which may occur at an early training stage.[344] Moreover, reaching the training compute threshold will likely be the primary trigger for the notification obligation under Article 52(1)'s first sentence.[345]

---

[336] AI Act, recital 111, fifth sentence; for a comprehensive discussion of the role of compute thresholds for AI governance see Matteo Pistillo and others, 'The Role of Compute Thresholds for AI Governance' (2025) 1 George Washington Journal of Law & Technology 26; for a discussion whether article 51(2) may lead to a 'downsizing effect' where providers use less training compute to not exceed the threshold, see Schneider and Schneider, 'Art. 51' (n 4) paras 34–35; also see, on the value of training compute as an approximation of model capabilities, Erben and others (n 80) 8; see also Somala, Ho and Krier (n 313) who argue that *pre-training* compute is becoming a less reliable proxy for a model's capabilities but also conclude that compute-based policies 'still offer key advantages that many other approaches lack'; for a critique of the training compute threshold, see Wachter (n 2) 697–698, 715.

[337] AI Act, art 51(2); article 3(67) defines a FLOP as 'any mathematical operation or assignment involving floating-point numbers, which are a subset of the real numbers typically represented on computers by an integer of fixed precision scaled by an integer exponent of a fixed base'.

[338] Martin Ebers, 'Truly Risk-Based Regulation of Artificial Intelligence How to Implement the EU's AI Act' (2025) 16 European Journal of Risk Regulation 684, 699; see also Martini (n 2) para 194.

[339] Robi Rahman and others, 'Over 30 AI models have been trained at the scale of GPT-4' (*Epoch AI*, 2025) <https://epoch.ai/data-insights/models-over-1e25-flop> accessed 7 January 2026.

[340] Jaime Sevilla and Edu Roldán, 'Training Compute of Frontier AI Models Grows by 4-5x per Year' (*Epoch AI*, 2024) <https://epoch.ai/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year> accessed 7 January 2026. However, it appears difficult to reliably predict the amount of training compute that will be spent on future GPAI models. Notably, Edelman and others, 'Why GPT-5 used less training compute than GPT-4.5 (but GPT-6 probably won't)' (*Epoch AI*, 2025) <https://epoch.ai/gradient-updates/why-gpt5-used-less-training-compute-than-gpt45-but-gpt6-probably-wont> accessed 7 January 2026, argued that OpenAI's GPT-5 used less training compute than GPT-4.5 but that this trend is not likely to continue.

[341] Carey (n 182) 9; see also Marco Almada and Nicolas Petit, 'The EU AI Act: Between the Rock of Product Safety and the Hard Place of Fundamental Rights' (2025) 62 Common Market Law Review 85, 102, who estimate that the scope of the AI Act's rules for GPAI models with systemic risk will likely be outdated soon; further see para 3.

[342] See Section 2.3.

[343] See Section 2.1.1.1.; see also Section 2.2.2.

[344] For example, a GPAI model that is trained with a cumulative amount of computation of 10^{26} FLOPs would reach the threshold under article 51(2) already at a point in time where only one tenth of the total amount of computation used for its training is spent, as the threshold under article 51(2) is 10^{25} FLOPs and 10^{26} FLOPs = 10×10^{25} FLOPs.

[345] See commentary on Article 52, Section 2.1.1.2.1. in this work.

90. Compute thresholds are not only relevant to the classification of GPAI models as presenting systemic risk. The Commission Guidelines envisage the use of a lower compute threshold to determine whether an AI model can be considered a GPAI model according to the definition under Article 3(63).[346] They indicate that, if a model is trained with more than 10^{23} FLOPs and can generate language (either textually or through audio), images from text, or videos from text, this indicates that it can be considered a GPAI model.[347] Training compute thresholds have also been explored as a means of AI regulation in the United States.[348]

2.2.1. Cumulative amount of computation used for training

91. Article 51(2)'s presumption applies where 'the cumulative amount of computation used for [the GPAI model's] training' exceeds 10^{25} FLOPs. It is clear from this wording ('training') that the computation used for the model's deployment, so-called inference compute,[349] does not play a role for the presumption of high-impact capabilities.[350] However, as different activities and methods involving computation ("computational activities") can play a role in the development of GPAI models,[351] a key question is which kinds of computational activities may be included in Article 51(2)'s compute count.

2.2.1.1. General rule

92. The AI Act does not define the training of a GPAI model. However, Recital 111 sets out that '[t]he cumulative amount of computation used for training includes the computation used across the activities and methods that are intended to enhance the capabilities of the model prior to deployment, such as pre-training, synthetic data generation and fine-tuning'.[352] Drawing on this recital, the Commission Guidelines and some legal scholars have taken the view that, '*as a general rule*, [...] *all [computational activities] that contributed or will contribute to the model's capabilities*'

[346] Commission Guidelines (n 39) para 17; see forthcoming commentary on Article 3(63) in this work.

[347] Commission Guidelines (n 39) para 17. One may note that the Commission Guidelines (n 39) para 115 define 'training compute' more narrowly in the context of article 3(63) than in the context of article 51(2). Further, see forthcoming commentary on Article 3(63) in this work.

[348] See California Senate Bill No. 53, Transparency in Frontier Artificial Intelligence Act, Sec 22757.11(i) [2025] and the rescinded U.S. Executive Order 14110 of October 2023 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, [2023] 88 F.R. 75191.

[349] See Code of Practice Safety and Security Chapter (n 22) Measure 1.2, point (1)(a) and Measure 7.6, point (2); Pablo Villalobos and David Atkinson, 'Trading Off Compute in Training and Inference' (*Epoch AI*, 2023) <https://epoch.ai/publications/trading-off-compute-in-training-and-inference> accessed 7 January 2026.

[350] See Schneider and Schneider, 'Art. 51' (n 4) 31. However, the foreseeable amount of inference compute may play a role for the obligation to assess and mitigate potential systemic risks under article 55(1)(b), as evidenced by Code of Practice Safety and Security Chapter (n 22) recital b which states that 'the Signatories also recognise that the assessment and mitigation of systemic risks should include, as reasonably foreseeable, [...] the computing resources available at inference time because of their importance to the model's effects, for example by affecting the effectiveness of safety and security mitigations.')

[351] For an overview of the development of general-purpose AI, including different training stages of a GPAI model see Bengio and others, International AI Safety Report (n 214) 30–36; further see Somala, Ho and Krier (n 313) s 2. This is also acknowledged by the Commission Guidelines (n 39) para 116.

[352] AI Act, recital 111, sixth sentence.

need to be taken into account.[353] The Commission Guidelines specifically mention that computational activities directly contributing to parameter updates should be included.[354]

93. A 2025 report prepared for the Commission's Joint Research Centre proposed another approach to whether a computational activity is included in Article 51(2)'s training compute count based on two criteria.[355] It proposed to take into account *only* computational activities that either *directly update the final model's parameters* or *create model-specific inputs* that depend on the current model state.[356] This approach has its merits since – as pointed out by the report – direct parameter updates 'form the core of model training and are unambiguously responsible for enhancing model capabilities'.[357] It is questionable, however, whether this consideration suffices for *generally*[358] excluding computational activities that influence the final model's weights only indirectly and do not create model-specific inputs. The distinction drawn between core and non-core computational activities finds no basis in the wording of Article 51(2), which does not distinguish between different computational activities performed for a model's training and therefore suggests that all such computational activities need to be taken into account. Moreover, Article 51(2) refers to the '*cumulative* amount of computation used for [the GPAI model's] training' (emphasis added). While this could be an acknowledgement of the evident fact that compute expenditure needs to be accumulated over time, it may also suggest a broader conception of training compute.[359] The same

[353] Commission Guidelines (n 39) para 116 (quoted, emphasis added). In the same sense: Bernsteiner and Schmitt, 'Art. 51' (n 6) para 37; Feiler, Forgó and Nebel, 'Article 51' (n 287) para 10. The Commission Guidelines, while not binding for providers of GPAI models (Commission Guidelines (n 39) para 9; for the non-binding nature of 'guidance documents' issued by the Commission see Case C-308/11 *Chemische Fabrik Kreussler & Co. GmbH v Sunstar Deutschland GmbH* [2012] EU:C:2012:548 paras 23–24), set out the Commission's interpretation of the AI Act (Commission Guidelines (n 39) para 9). For a discussion of exceptions from this general rule, including under the Commission Guidelines, see Section 2.2.1.2.

[354] Commission Guidelines (n 39) para 116. The Commission Guidelines (n 39) para 118 further mention that in cases where a GPAI model has been created by combining model weights – including techniques like model weight merging, model weight averaging or integration of pre-existing model weights – 'the training compute used to train the combined model weights should be included in the estimation of the cumulative training compute of the model'. This follows from the Commission Guidelines' general rule of taking all computational activities contributing to the model's capabilities into account (see Commission Guidelines (n 39) para 116).

[355] Erben and others (n 80)

[356] Erben and others (n 80) 21–22. The Commission Guidelines (n 39) para 116 highlight that 'compute directly contributing to parameter updates' counts towards article 51(2)'s training compute threshold without establishing the direct contribution to a final model's parameters as an independent criterion in the context of article 51(2). The Commission Guidelines (n 39) para 115 draw an express distinction between two notions of 'training compute', one relating to the assessment of whether a GPAI model meets article 51(2)'s threshold and one for GPAI models that are not GPAI models with systemic risk. The Commission Guidelines establish a determinative criterion of direct contribution only for the latter purpose (see Commission Guidelines (n 39) para 115: 'In these guidelines, "training compute" of a general-purpose AI model refers to either: • the total amount of compute directly contributing to parameter updates in the model if the model is not a general-purpose AI model with systemic risk; • the cumulative amount of compute used to train the model if, and for the purpose of assessing whether, the model is a general-purpose AI model with systemic risk.'). By contrast, for the former purpose, 'compute directly contributing to parameter updates in the model' serves merely as an example of compute that should count towards article 51(2)'s compute threshold.

[357] Erben and others (n 80) 22.

[358] For a discussion of whether this criterion may justify exceptions for specific computational activities, see Section 2.2.1.2.

[359] See Pistillo and others (n 336) 33: 'Estimates of "training compute" typically refer only to the amount of compute used during pretraining. More specifically, they refer to the amount of compute used during the final pre-training run, which contributes to the final machine learning model, and does not include any previous test runs or post-training enhancements, such as fine-tuning. There are exceptions: for instance, the EU AI Act considers the *cumulative* amount of compute used for training by including all the compute "used across the activities and methods

applies to Recital 111, which expressly mentions synthetic data generation as a computational activity that does not lead to direct parameter updates.[360] Moreover, there is no generally agreed upon definition of an AI model's training that excludes computational activities indirectly updating the final model's parameters.[361] Introducing an additional criterion such as whether the computational activity leads to direct parameter updates may come with an increased risk of regulatory arbitrage where providers may seek to circumvent Article 51(2) by interposing an additional step between a compute-intense activity and the parameter updates.[362]

94. Similarly, the mere fact that model-dependent processes 'generate training signals [...] *specifically tailored* to the model's capabilities or are inseparable from the capability development process' (emphasis added) and 'cannot be trivially reused for other training pipelines' – the rationale presented for including model-dependent processes[363] – does not appear sufficient justification for excluding model-independent processes that enhance a model's capabilities from Article 51(2)'s compute count.[364] Article 51(2)'s text does not differentiate between model-dependent and model-independent processes. Recital 111 specifies that Article 51(2) includes computational activities that are '*intended* to enhance the capabilities of the model prior to deployment' (emphasis added). While this criterion could be interpreted in a way that comes close to the concept of 'specifically tailored' training signals,[365] Recital 111 simultaneously cites synthetic data generation as an example of a computational activity that need not be model-dependent but should nevertheless be included

---

that are intended to enhance the capabilities of the model prior to deployment, such as pre-training, synthetic data generation and fine-tuning."' (emphasis by authors).

[360] The report assumes, however, that synthetic data generation may in certain instances be covered by article 51(2)'s training compute threshold where it serves as model-specific input (see Erben and others (n 80) 25–26, 49). For a discussion of the role of synthetic data generation in the context of article 51(2) see Section 2.2.1.2.1.

[361] In general, there appears to be no consensus as to a definition of the training of an AI model. Heim and Koessler (n 227) 7 describe the training of an AI model as 'an iterative process where a model—a large amount of numeric values (the so-called "parameters") arranged in a certain way (the so-called "architecture")—is exposed to a large amount of data, allowing the model to learn from the data by adapting the parameters.' It is uncertain to what extent one can distinguish between the training and the development of an AI model. Both notions are sometimes used interchangeably (see, for example, Erben and others (n 80) 15: 'Training compute—the computational resources invested in developing an AI model, [...].'). The increasing complexity of frontier AI model training pipelines (see Somala, Ho and Krier (n 313) s 2) makes it challenging to define a model's training in a future-proof way.

[362] See, however, Erben and others (n 80) 9 who consider that '[e]ach regulatory definition [for "cumulative compute" in the EU AI Act] established today inadvertently creates incentives for architectures that comply technically while circumventing the spirit of compute accounting, with the potential unintended consequence of transforming "cumulative compute" into a perpetual regulatory challenge requiring frequent reassessment.'; for an overview over different ways in which provides may seek to circumvent training compute thresholds, see Matteo Pistillo and Pablo Villalobos, 'Defending Compute Thresholds Against Legal Loopholes' (2025) <https://arxiv.org/abs/2502.00003> accessed 7 January 2026 2–3.

[363] Erben and others (n 80) 22.

[364] Opposing view: Erben and others (n 80) 22.

[365] This would be the case if this criterion were understood as excluding computational activities originally performed for a different purpose but subsequently repurposed to serve the model's training. A different interpretation, however, of recital 111's reference to computational activities 'intended to enhance the capabilities of the model prior to deployment' appears more persuasive. As article 51(2)'s training compute threshold is linked to a presumption of high-impact capabilities, the legislature likely intended to indicate that computational activities performed during the model's training for the purpose of enhancing the model's security should not count towards the compute threshold. (see Commission Guidelines (n 39) para 119 that specify that 'compute spent on purely diagnostic processes that do not contribute to enhancing model capabilities, such as model evaluations or red-teaming' needs not be included in the compute count). In this context, the reference to intended capability enhancements, rather than actual enhancements, is plausibly intended to preclude arguments over whether a particular computational activity actually resulted in capability enhancements – a matter that may often prove difficult to establish in practice.

in the cumulative amount of compute used for training under Article 51(2).[366] Moreover, the criterion of model-dependency may not be able to fully account for the increasing complexity of model training pipelines where different models may play a role in the process of training a GPAI model.[367] For example, where training includes the training of different models that are later combined into one model,[368] it may be unclear to which model(s) the model-dependency criterion applies. It is also uncertain how the criterion relates to the AI Act's understanding of a model's lifecycle,[369] which may imply that different models used in training – such as parent models used in distillation[370] – coincide with the final model from the AI Act's legal point of view.[371] Finally, the introduction of this additional criterion could give rise to a greater risk of gamification as well.

2.2.1.2. Exceptions

95. Proceeding on the basis of the general rule that, in principle, all computational activities contributing to the model's capabilities are to be included in Article 51(2)'s training compute count,[372] the question arises of whether exceptions exist to this general rule. The Commission Guidelines suggest the existence of such exceptions and list six examples of computational activities that, seemingly in contrast to the general rule,[373] do not need to be counted towards Article 51(2)'s training compute threshold, while highlighting that this list may be subject to change due to technological developments.[374] The listed examples are (i) the generation of publicly accessible synthetic data,[375] (ii) 'purely diagnostic processes' not contributing to the model's capabilities (such as model evaluations), (iii) computational activities which 'contribut[e] to enhancing model capabilities only through lessons learnt by humans' (such as exploratory research projects or failed experiments in

[366] For a discussion of synthetic data generation, see Section 2.2.1.2.1.

[367] See Somala, Ho and Krier (n 313) s 2; see also Pistillo and Villalobos (n 362) 14–15.

[368] See Somala, Ho and Krier (n 313) s 2; Frontier Model Forum, 'Issue Brief: Measuring Training Compute' (2024) <https://www.frontiermodelforum.org/updates/issue-brief-measuring-training-compute/> accessed 7 January 2026 (The Frontier Model Forum is an organisation supported by frontier AI model providers, see https://www.frontiermodelforum.org/about-us/ accessed 7 January 2026. For an overview over model merging methods used in training AI models see Yang and others, 'Model Merging in LLMs, MLLMs, and Beyond: Methods, Theories, Applications and Opportunities' (2024) <https://arxiv.org/abs/2408.07666> accessed 7 January 2026; see also Commission Guidelines (n 39), para 118.

[369] See AI Act, recital 113, third sentence, and recital 114, third sentence. For the Commission's interpretation of a model's lifecycle, see Commission Guidelines (n 39) para 21–22.

[370] For a discussion of knowledge distillation, see Section 2.2.1.2.2.

[371] See Commission Guidelines (n 39) para 119. Further, see Frontier Model Forum (n 368) arguing in favour of considering different independently-trained AI models as a single model for the purpose of applying compute thresholds under certain circumstances ('Some AI models are produced by taking multiple independently-trained AI models, combining them, and then further training the result to integrate them – and these should be considered an individual model, with all the resulting computation. Other composite AI systems may operate by sampling from a variety of underlying models that have been trained separately but have never been jointly-trained (such as a chat model and a separate safety-filtering model). Though these are part of a system, they should not be considered part of the same "model" and only the compute of each individual model should be reported (if necessary).')

[372] See Section 2.2.1.1.

[373] The Commission Guidelines introduce the general rule in paragraph 116 and, following remarks on synthetic data generation and model weight combination, set out these examples in paragraph 119, with the language indicating that these examples contrast with the general rule (see Commission Guidelines (n 39) para 119: 'By contrast, the following are examples of compute which need not be included in the estimation of the cumulative training compute. This list may change as technology evolves: [...]').

[374] Commission Guidelines (n 39) para 119.

[375] For a discussion of this exception, see Section 2.2.1.1.1.

synthetic data generation),[376] (iv) the training of parent models used in distillation,[377] (v) the training of auxiliary models (such as reward models) and (vi) recomputation of activations to save memory.[378]

96. It is interesting to consider the extent to which, and the grounds upon which, the exclusion of these examples and other potential exceptions can be justified. The second and sixth examples – purely diagnostic processes and recomputations of activations – appear most straightforward to justify, as they rather represent a confirmation of the general rule. Purely diagnostic processes, by definition, do not contribute to the model's capabilities.[379] Recomputation of activations has been developed as a technique to address memory limitations of the hardware used in AI model training.[380] To avoid failure of model training due to running out of memory, some of the activations, which may be described as intermediate computational results generated during the model's training,[381] are not stored but discarded and later regenerated.[382] As these recomputations in essence 'trad[e] less memory for greater computation',[383] they can also be regarded as not resulting in an enhancement of a model's capabilities.[384]

97. The other examples listed in the Commission Guidelines appear less straightforward on the basis of the Act's text. In particular, special attention must be paid to the particular cases of synthetic data generation and knowledge distillation, which are discussed in more detail in subsequent sections.[385] In general, when considering the scope of what constitutes 'cumulative training compute', it is necessary to consider Article 51(2)'s integration within the systematic context of the classification rules in Section 1 of Chapter V. Many concerns that could, in principle, support an exception for certain types of compute appear to be already addressed to some extent by the AI Act's provision for a procedure to contest the training compute-based classification under Article 52(2) and (3), which allows providers to rebut Article 51(2)'s high-impact capabilities presumption.[386] This applies, for example, in cases where a computational activity only marginally contributes to the enhancement of a model's capabilities or has a particularly low compute-to-capability ratio, as such considerations could serve to rebut the high-impact capabilities presumption. Moreover, far-reaching exceptions from the training compute count may also prove unnecessary in light of Article 51(3), which permits the Commission to adjust the compute threshold, including the computational activities counting towards the threshold, via a delegated act where necessary.[387]

---

[376] Similar for training compute thresholds in general: Frontier Model Forum (n 368) ('Discarded versions or branches should not be included. Model developers will often experiment with different branches and versions of a given model that are ultimately discarded. Since information from such branches is not explicitly included in the final model, the operations used to train discarded branches should not be included in measures of training compute.')

[377] For a discussion of this exception see Section 2.2.1.1.2.

[378] Commission Guidelines (n 39) para 119; to the same effect for training compute thresholds in general: Frontier Model Forum (n 368).

[379] See Commission Guidelines (n 39) para 119.

[380] Ping Chen and others, 'Optimizing Large Model Training through Overlapped Activation Recomputation' (2025) <https://arxiv.org/abs/2406.08756> accessed 7 January 2026 1.

[381] Activations are computed with the help of so-called activation functions which play an important role in a neural network's ability to learn from data. For an introduction see Niklas Lang, 'Activation Functions in Neural Networks: How to Choose the Right One' (*Towards Data Science*, 2024) <https://towardsdatascience.com/activation-functions-in-neural-networks-how-to-choose-the-right-one-cb20414c04e5/> accessed 7 January 2026.

[382] Ping and others (n 380); Frontier Model Forum (n 368).

[383] Frontier Model Forum (n 368).

[384] Frontier Model Forum (n 368).

[385] For synthetic data generation, see Section 2.2.1.2.1.; for knowledge distillation, see Section 2.2.1.2.2.

[386] See commentary on Article 52, Section 2.2.2.1. in this work.

[387] See Section 2.3.1.

2.2.1.2.1. Synthetic data generation

98. Synthetic data – which is not directly based on observation of the real world but rather artificially generated from real data[388] – plays an increasingly important role in the training of AI models.[389] Microsoft's phi-4 model exemplifies this trend.[390] Around twenty-five percent of the model's pre-training compute budget was spent on the generation of synthetic data.[391] While only having 14 billion parameters and therefore being small in comparison with frontier AI models, phi-4 scored better on the mathematics and science benchmarks MATH Level 5 and GPQA Diamond than did OpenAI's GPT-4o – the model that was used to generate the synthetic data for phi-4's training.[392] The Commission Guidelines specify that the generation of *non-publicly accessible* synthetic data should be included in Article 51(2)'s compute count,[393] while mentioning the generation of *publicly accessible* synthetic data as an example of a computational activity that should be excluded because it 'may be indistinguishable from other publicly accessible data'.[394] Some authors have even argued that the generation of synthetic data should not be included regardless of whether it is publicly available, contending in particular that synthetic data generation precedes rather than forms part of the model's training.[395]

99. Several arguments support these views. While no apparent consensus has emerged as to a definition of a GPAI model's training, there are definitions of model training which appear to exclude synthetic data generation.[396] Further, synthetic data serves as a functional replacement of real data whose generation is also not included in Article 51(2)'s compute count.[397] Moreover, it appears possible that in certain instances – particularly where a provider acquires synthetic data from a third party – it is difficult to know whether the data has been synthetically generated and the amount of compute spent to generate the data.[398]

---

[388] Synthetic data has been defined as 'artificial data that is generated from original data and a model that is trained to reproduce the characteristics and structure of the original data' (European Data Protection Supervisor, 'Synthetic Data' <https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en> accessed 7 January 2026) and contrasted with 'real data, which is generated not by a model but by real world systems' (James Jordon and others, 'Synthetic Data – What, Why and How?' (2022) <http://arxiv.org/abs/2205.03257> accessed 7 January 2026, 5). Research has demonstrated that the use of synthetic data in training AI models can improve capabilities, at least within certain domains (Pablo Villalobos and others, 'Will We Run Out of Data? Limits of LLM Scaling Based on Human-Generated Data' (*Epoch AI*, 2024) <https://epoch.ai/blog/will-we-run-out-of-data-limits-of-llm-scaling-based-on-human-generated-data> accessed 7 January 2026).

[389] See also Villalobos and others (n 388).

[390] Somala, Ho and Krier (n 313) s 2.

[391] Somala, Ho and Krier (n 313) s 2.

[392] Jean-Stanislas Denain, 'Models With Downloadable Weights Currently Lag Behind the Top-Performing Models' (*Epoch AI*, 2024) <https://epoch.ai/data-insights/open-vs-closed-model-performance> accessed 7 January 2026; Somala, Ho and Krier (n 313) s 2.

[393] Commission Guidelines (n 39) para 117 ('If the model is trained on synthetic data that is not publicly accessible, the forward passes used to generate the data, including discarded data, should be included in the estimation of the cumulative training compute. For example, if 100 samples were generated and only the top 10 samples were selected for training, the compute used to generate all 100 samples should be counted since the compute used to generate all 100 samples was necessary to create the selected 10.')

[394] Commission Guidelines (n 39) para 119.

[395] Haar and Siglmüller, 'Art. 51' (n 9) para 41; see also Erben and others (n 80) 21–22, 25–26, 49 who do not reject the inclusion of compute spent on synthetic data generation from the outset but establish a framework that 'effectively negates the question of whether to account for general-purpose synthetic data at all'.

[396] This is presupposed by Haar and Siglmüller, 'Art. 51' (n 9) para 41.

[397] Haar and Siglmüller, 'Art. 51' (n 9) para 41.

[398] See Commission Guidelines (n 39) para 119; Haar and Siglmüller, 'Art. 51' (n 9) para 41.

100. However, there are also good reasons to include synthetic data generation.[399] The inclusion of compute spent on synthetic data generation in a model's pre-training compute budget by some authors suggests that including synthetic data generation within the notion of training does not exceed the wording of Article 51(2).[400] This is reinforced by Recital 111 specifically mentioning synthetic data generation among the activities and methods to be included in Article 51(2)'s compute count, regardless of its public availability.[401]

101. Furthermore, in many cases the provider should be able to estimate the amount of compute spent on synthetic data generation – for instance, by tracking compute expenditure where it generates the data itself or through contractual arrangements governing data purchases.[402] Indeed, Article 53(1)(a) requires providers to possess at least some information about synthetic data generation, which could serve as the basis for such estimates. This provision requires providers of GPAI models to draw up technical documentation – including information on the data used for training, in particular the type and provenance of data and how it was obtained, as well as information on the computational resources used to train the model[403] – for the purpose of providing it, upon request, to the AI Office and the national competent authorities.[404] This includes information on the use and provenance of synthetic data for the model's training.[405]

102. This documentation requirement may limit the risk of regulatory arbitrage associated with including compute spent on synthetic data generation in Article 51(2)'s compute count, as providers that seek to generate or acquire synthetic data in a manner that makes it impossible for them to estimate the compute spent on its generation may risk breaching their obligation under Article 53(1)(a).[406] Accordingly, cases where it proves genuinely impossible for the provider to produce even an estimate of the compute spent on synthetic data generation may be edge cases that do not necessarily warrant the general exclusion of synthetic data generation.

103. Overall, the more compelling arguments support including synthetic data generation in Article 51(2)'s compute count, in closer alignment with the Commission Guidelines' general rule to include all computational activities that contribute to the model's capabilities.[407] This approach is further supported by the consideration that an additional differentiation – for example, between publicly and non-publicly accessible synthetic data[408] – raises new demarcation questions, such as whether and under which circumstances it suffices for a provider to offer synthetically generated data for sale in order for it to be considered publicly accessible. Moreover, cases where the compute threshold is surpassed due to synthetic data generation without significant capability enhancement can be addressed through the procedure to contest classification set out in Article 52(2) and (3).[409]

---

[399] In favour of inclusion of synthetic data generation: Carey (n 182) 9; Somala, Ho and Krier (n 313) s 2; see also Feiler, Forgó and Nebel, 'Article 51' (n 287) para 10.

[400] See Somala, Ho and Krier (n 313) s 2.

[401] AI Act, recital 111, sixth sentence. For the relevance of the recitals for the purposes of interpretation see *Puppinck* (n 12) paras 75–76.

[402] Critical of the viability of such contractual arrangements: Haar and Siglmüller, 'Art. 51' (n 9) para 41.

[403] AI Act, art 53(1)(a) in conjunction with AI Act, annex XI, s 1, points 2(c) and (d).

[404] For a discussion of this requirement see commentary on Article 53, Section 2.1.1. in this work.

[405] See commentary on Article 53, para 24 in this work.

[406] The extent to which article 53(1)(a) limits the risk of regulatory arbitrage depends on the extent of the obligation under article 53(1)(a) which is discussed in commentary on Article 53, para 34 in this work.

[407] To the same effect: Somala, Ho and Krier (n 313) s 2 ('[T]he EU AI Act does include synthetic data generation in training compute [...]'); Carey (n 182) 9; see Section 2.2.1.1.

[408] See Commission Guidelines (n 39) paras 117, 119.

[409] See also Section 2.2.1.1.

2.2.1.2.2. Knowledge distillation

104. (Knowledge) distillation is a method for model training that involves a smaller "student model" and a larger "teacher model".[410] Rather than undergoing standard pre-training, the student model is trained to "mimic" the output of the teacher model.[411] By "reusing" the teacher model,[412] smaller student models require fewer computational resources and operate faster than their teacher models, without necessarily sacrificing capabilities.[413] For example, researchers successfully distilled Google's BERT model into the DistilBERT model, using only 3% of the computational resources required for training the original model while retaining 97% of BERT's language understanding capabilities and achieving 60% faster performance.[414]

105. Three computational activities associated with distillation warrant consideration under Article 51(2): first, the training of the teacher model; second, the generation of the teacher model's outputs that serve as a basis for the training of the student model; and thirdly, the training of the student model with this output.[415] As with synthetic data generation, interpretive uncertainty arises as to whether some or all of these computational activities must be taken into account.

106. Under the Commission Guidelines' general rule laid out above,[416] all three computational activities would be included in the student model's compute count under Article 51(2) as they all, at least indirectly, contribute to the student model's capabilities.[417] However, there are arguments for an exception from this rule, in particular with regard to the first computational activity, the training of the teacher model.[418] One could contend that the amount of computation used for training a student model's teacher model is not 'used for *its* [that is, the student model's] training' (emphasis added) in the sense of Article 51(2). Moreover, where a teacher model serves purposes beyond distillation, one may argue that the training of the teacher model is not '*intended* to enhance the capabilities of the [student] model' in the sense of Recital 111's sixth sentence but rather serves the training of the teacher model itself.[419]

107. While these arguments carry some weight, they do not appear entirely convincing. As evidenced by Recital 111's reference to synthetic data generation, there are good reasons to assume that a computational activity is not required to directly contribute to the (student) model's capabilities to be included in Article 51(2)'s compute count.[420] Additionally, it is questionable whether the fact that a computational activity serves an additional purpose distinct from contributing to the (student) model's capabilities is a sufficient reason for exclusion. By its wording, Article 51(2) does not require a computational activity to only serve the model's training in order to be included. This is reinforced by a purposive argument that such an exclusivity requirement would lead to greater delineation

[410] Victor Sanh and others, 'DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter' (2020) <https://arxiv.org/abs/1910.01108v4> accessed 7 January 2026 2; Somala, Ho and Krier (n 313) s 2.
[411] Somala, Ho and Krier (n 313) s 2.
[412] See Pistillo and Villalobos (n 362) 14–18 for a discussion how 'model reuse' techniques such as knowledge distillation, kickstarting and reincarnation could be used as a loophole for training compute thresholds.
[413] Somala, Ho and Krier (n 313) s 2.
[414] Sanh and others (n 410); Somala, Ho and Krier (n 313) s 2.
[415] Somala, Ho and Krier (n 313) s 2.
[416] See Section 2.2.1.1.
[417] See Section 2.2.1.1.
[418] For a general discussion of exceptions from the Commission Guidelines' general rule, see Section 2.2.1.2.
[419] For different conceivable interpretations of recital 111's notion of 'activities and methods that are intended to enhance the capabilities of the model prior to deployment', see Section 2.2.1.1., n 362.
[420] See also Section 2.2.1.1.

difficulties and loopholes as a provider could seek to avoid the inclusion of a computational activity by assigning a secondary purpose to it.[421]

108. Considering whether and under which circumstances the teacher model and the student model may be regarded as one and the same model under the AI Act adds an additional layer of complexity to the question of whether the teacher model's training compute should be included in the student model's compute count. There appears to be no simple answer to the relationship between a teacher model and a student model under the AI Act, which relates to the broader question of how the AI Act treats modification and subsequent development of a model.[422] The Commission Guidelines assume that where the same provider provides both the teacher model and the student model, '[a]ny subsequent development of the model downstream of [the model's] large pre-training run [...] forms part of the same model's lifecycle rather than giving rise to new models' while also acknowledging the difficulty of delineating a model and its lifecycle.[423] If one were to follow this and further regard distillation as a 'subsequent development of the model', a teacher model and its student model could be one and the same under the AI Act, likely implying that the teacher model's training compute would need to be taken into account. A purposive argument in favour of this interpretation is that it could avoid a potential loophole described in literature: a provider could develop a teacher model above the training compute threshold, use distillation to train a student model below the threshold, and only place the student model on the market.[424] However, such an interpretation appears far from certain, and the Commission Guidelines' approach only relates to the case where the same provider provides both the original and modified model.[425]

109. To further add to the complexity, one can inquire about which kind of allocation of responsibilities between a provider of a teacher model and a provider of a student model with regard to the obligations under Article 55 stemming from systemic risk classification[426] would be most faithful to the purpose of these obligations. It appears that certain systemic risks can be most effectively, or even only, addressed by the provider of the teacher model, while others are better addressed by the student model provider.[427] Moreover, a categorical focus on only one of the providers could be insufficient for the effective mitigation of systemic risks where one of the models is not placed on the Union market and is therefore outside the scope of the AI Act.[428] Conversely, arguments against

[421] For example, a provider could use synthetically generated data to train a second model to prevent the compute spent on the generation of this synthetic data from being counted.

[422] For an analysis on the regulation of GPAI model modifications under the AI Act, see forthcoming commentary on Modifications in this work.

[423] Commission Guidelines (n 39) para 22.

[424] See Pistillo and Villalobos (n 362) 18.

[425] See Commission Guidelines (n 39) para 22; see also the forthcoming chapter on Modifications in this work.

[426] For a discussion of these obligations see commentary on Article 55 in this work.

[427] For example, systemic risks stemming from data poisoning appear better mitigated by the provider of the teacher model as it controls the teacher model's pre-training data, which typically uses large amounts of data available on public web and is therefore prone to such attacks (for the notion of data poisoning and the poisoning attacks it enables see Alexandra Souly and others, 'Poisoning Attacks on LLMs Require a Near-Constant Number of Poison Samples' (2025) <https://arxiv.org/abs/2510.07192> accessed 7 January 2026 1; see also Code of Practice Safety and Security Chapter (n 22) Measure 5.1 which lists 'filtering and cleaning of training data' as an example of a safety mitigation). On the other hand, the student model provider appears better positioned to mitigate systemic risks which may arise following the student model's deployment and require the implementation of additional guardrails for deployment. This includes the risk of misuse of the student model for large-scale cyber-attacks by a malicious actor who bypassed the model's guardrails (for the risk of users bypassing an AI system's guardrails for harmful requests, see Bengio and others, International AI Safety Report (n 214) 197–198; see also Code of Practice Safety and Security Chapter (n 22) Measure 5.1 which lists 'fine-tuning the model to refuse certain requests' as an example of a safety mitigation).

[428] See AI Act, art 2(1)(a); see also forthcoming commentary on Article 2 in this work.

unnecessary double obligations for both the provider of the student model and the provider of the teacher model can be made, at least where this does not contribute to a more effective mitigation of systemic risks.[429]

2.2.1.3. Methods for determining the cumulative amount of computation

110. As there are different computational activities that may contribute to a model's cumulative amount of computation,[430] there are different conceivable methods for determining the amount of computation spent on each of these computational activities.[431] This observation is reflected in point (c) of Annex XIII, which lays out that the amount of computation used for a model's training can be both 'measured in floating point operations' or 'indicated by a combination of other variables', including estimates of training cost, duration or energy consumption.[432]

2.2.1.3.1. Estimation methods and overall accuracy of the estimate

111. The Commission Guidelines state that a provider 'may choose any method to estimate the relevant amount of training compute, so long as the estimated amount is, in the providers' best judgement, *accurate within an overall error margin of 30%* of the reported estimate.'[433] The use of some error margin appears appropriate given the difficulties in *precisely* measuring compute expenditure across all relevant computational activities.[434] However, a more dynamic permissible overall error margin instead of relying on a fixed overall error margin of 30% may be more appropriate in the context of Article 51(2). While a fixed error margin appears particularly sensible where a provider must determine computational resources independently of any specific threshold,[435] it may impose unnecessary estimation effort and produce difficult-to-justify results regarding the question of whether the training compute threshold has been crossed. Two examples may illustrate these difficulties:

---

[429] See AI Act, recital 109, third sentence: 'In the case of a modification or fine-tuning of a model, the obligations for providers of general-purpose AI models should be limited to that modification or fine-tuning, for example by complementing the already existing technical documentation with information on the modifications, including new training data sources, as a means to comply with the value chain obligations provided in this Regulation.'

[430] See Section 2.2.1.

[431] See Erben and others (n 80) 9 ('No standardised methodology exists for measuring training compute across different architectures and training paradigms [...].'). For example, the Commission Guidelines (n 39) para 117 provide specific guidance on how to measure the compute expenditure on the generation of non-publicly accessible synthetic data.

[432] Annex XIII is particularly relevant to classification under article 51(1)(b)'s classification condition (see Section 2.1.2.1.2.) but can also play a (more limited) role for article 51(1)(a) (see Section 2.4.2.), which article 51(2) relates to. One may further note that point (c) of annex XIII relates to 'the amount of computation', whereas article 51(2) relates to 'the *cumulative* amount of computation' (emphasis added).

[433] Commission Guidelines (n 39) para 120 (emphasis added).

[434] Beyond the Frontier Model Forum's general recommendations (Frontier Model Forum (n 368)), no industry standard for measuring training compute exists as of writing (December 2025) (see Pistillo and others (n 336) 60, 62; Heim and Koessler (n 227) 7). While training compute is generally regarded as readily measurable as 'it can be directly calculated from model specifications or inferred from data about the use of hardware with minimal effort' (see Heim and Koessler (n 227) 10; further, see Pistillo and others (n 336) 62), frontier AI model training is becoming increasingly complex (see Somala, Ho and Krier (n 313) s 2) and not all relevant computational activities may be equally straightforward to account for by a provider, in particular where they are performed by different actors – for example where the provider uses synthetic data generated by different party for the model's training.

[435] See AI Act, art 53(1)(a) in conjunction with annex XI, s 1, point 2(d) for the provider's obligation to determine the computational resources used to train the model and provide this information, upon request, to the AI Office and the national competent authorities.

112. First, suppose a method estimates a model's cumulative amount of computation at 10^{24} FLOPs with an *overall error margin of 50%*. This would mean that the model was trained with no more than 1.5×10^{24} FLOPs, compellingly proving it does not meet Article 51(2)'s 10^{25} FLOPs threshold. Yet the Commission Guidelines would deem this method impermissible because its error margin exceeds 30%. Second, suppose a method estimates a model's cumulative amount of computation to be 10^{25} FLOPs with an overall error margin of 30%. Depending on the positioning of the estimate within this error margin, there could be a substantial probability that the model exceeds the threshold. Yet the Commission Guidelines suggest this method would constitute sufficient proof that the model does not exceed Article 51(2)'s threshold.

113. These examples illustrate that setting a permissible overall error margin for reported estimates in the abstract may produce undesirable results. Moreover, it appears difficult to derive such a threshold from the AI Act itself, which provides no guidance on permissible error margins. Rather, a different approach, relying on a more dynamic error margin, appears worth considering: the closer a model's estimated computation is to Article 51(2)'s threshold, the greater the precision required of a method to prove whether the threshold is met or not.[436] Such an approach would come with its own limitations[437] but could be suitable to avoid the difficult-to-justify results laid out above.

114. An additional suggestion by some scholars that computational activities consuming less than 20% of the overall compute budget *cumulatively* should not be taken into account also has some practical appeal.[438] In many cases, activities accounting for only a small portion of the training compute budget may indeed be negligible. However, the proposal may lack necessary nuance where a model's training compute is close to the set threshold, as these computational activities could prove decisive in determining whether the training compute threshold is met.

2.2.1.3.2. Available methods

115. The Commission Guidelines lay out two possible approaches for providers to estimate training compute – a hardware-based approach and an architecture-based approach.[439] Under the hardware-based approach, the (expected) training compute is estimated based on the number of graphics processing units ("GPUs") or other hardware units used for training, the total duration of their use, their peak theoretical performance and their average percentage of utilisation.[440] This differs from the architecture-based approach, which estimates the (expected) training compute based on the number of full passes made during the training of a neural network and the total number of operations performed in a full pass.[441] Alternatively, according to the Commission Guidelines, for

[436] This approach aligns with the Frontier Model Forum (n 368)'s recommendation that for the application of compute thresholds – not related to the AI Act specifically – the required precision for training compute estimates should be context-dependent.

[437] In particular, requiring a lower error margin in instances where a model is close to Article 51(2)'s training compute threshold comes with increased costs for the provider and the Commission to determine and verify the compute expenditure for a model. Moreover, there are practical limits to the achievable precision for a compute estimate.

[438] This is suggested by Erben and others (n 80) 21–22. See further Frontier Model Forum (n 368) where the Frontier Model Forum recommends that for the application of compute thresholds – not related to the AI Act specifically – approximations that 'cumulatively change the total compute used by <5%' should be considered valid.

[439] Commission Guidelines (n 39) paras 121–130; see also Erben and others (n 80) 30–34.

[440] Commission Guidelines (n 39) para 123; see also Erben and others (n 80) 30–31 for the advantages and disadvantages of the hardware-based approach.

[441] Commission Guidelines (n 39) para 128 which note that this approach may only be used for GPAI models based on neural networks that are trained through a succession of forward and backward passes. One full pass is the combination of a forward pass and a backward pass; see also Erben and others (n 80) 31–32; Heim and Koessler (n 227) 10.

some models based on a dense transformer architecture, the (expected) training compute may be estimated based on the total number of model parameters and the total number of training tokens used for training.[442]

116. As the AI Act does not prescribe a particular method of determining whether a model exceeds Article 51(2)'s compute threshold, the approaches laid out in the Commission Guidelines appear *merely illustrative*, with other methods being admissible as well.[443] The Commission Guidelines, by their nature, cannot bind GPAI model providers.[444]

117. Moreover, an obligation to determine the cumulative amount of computation used for the model's training via a 'dual estimation methodology' using *both* a hardware-based approach and the architecture-based approach, suggested by some scholars,[445] does not appear supported by the AI Act or the Commission Guidelines.[446] Further methods that could play a role in determining or verifying whether Article 51(2)'s threshold is met could include inferences from the cost, duration or energy consumption of training;[447] inferences from benchmark performance;[448] or testimony from whistleblowers.[449]

2.2.1.3.3. Additional considerations

118. All operations should be counted equally, independent of floating-point precision (e.g. FP8, FP16, FP32[450]).[451] By its wording, Article 51(2) is concerned with the number of FLOPs, not with the precision of the number format used in these operations. This interpretation is further reinforced by the purpose of Article 51(2)'s threshold, which serves as an indicator for a model's high-impact capabilities, since floating-point precision appears to be only marginally indicative of a model's capabilities.[452]

---

[442] Commission Guidelines (n 39) para 129.

[443] The Commission Guidelines (n 39) para 120 state in that regard that providers 'may choose *any method* to estimate the relevant amount of training compute' (emphasis added) that meets the error margin requirement discussed above (see Section 2.2.1.3.2.).

[444] See Commission Guidelines (n 39) para 9: 'These guidelines are not binding for providers of general-purpose AI models; an authoritative interpretation of the AI Act may only be given by the Court of Justice of the European Union ("CJEU"). Nevertheless, these guidelines set out the Commission's interpretation and application of the AI Act, on which it will base its enforcement action.'; further, see *Chemische Fabrik Kreussler* (n 353) paras 23–24.

[445] See Erben and others (n 80) 21–22.

[446] See Commission Guidelines (n 39) para 121: 'Providers may choose to estimate the relevant amount of training compute by tracking graphics processing unit ('GPU') usage (hardware-based approach) *or* by estimating operations directly based on the relevant model's architecture (architecture-based approach), as appropriate to what is being estimated.' (emphasis added).

[447] See AI Act, annex XIII, point (c); for inferral from energy consumption, see Erben and others (n 80) 43.

[448] See Erben and others (n 80) 46.

[449] See Erben and others (n 80) 46.

[450] 'FP' is short for floating point and the number (8, 16, 32, ...) represents the number of bits associated with a number format. A higher number of bits may lead to higher accuracy during model training but may be associated with slower computation and higher memory usage. For an accessible explanation of different number formats used in AI development see James Chiang, 'Two Things You Should Know as an AI Beginner' (*Medium*, 2024) <https://medium.com/@tsunhanchiang/two-things-you-should-know-as-an-ai-beginner-4c4c011ff06a> accessed 7 January 2026.

[451] Commission Guidelines (n 39) para 122; Frontier Model Forum (n 368).

[452] See Erben and others (n 80) 27; Frontier Model Forum (n 368); see Erben and others (n 80) 27, 30 for the motivations behind using different number formats.

119. According to its Guidelines, the Commission expects providers 'to document the assumptions made in making their estimations, including the method of estimation, and the associated uncertainties.'[453] Such a documentation requirement implicitly follows from Article 52(1)'s second sentence as well. Where a model meets Article 51(2)'s training compute threshold, its provider is not only required to notify the Commission of this fact but also to provide the information necessary to demonstrate that this is the case.[454] While there is uncertainty as to the extent of information a provider must submit, this includes information about the method of estimation.[455]

2.2.2. Effect of the presumption and rebuttability

120. Where a model meets Article 51(2)'s training compute threshold, it is presumed to have high-impact capabilities pursuant to Article 51(1)(a)[456] and is thus automatically classified as presenting systemic risk.[457] Via Article 51(1)(a), this high-impact capabilities presumption is further linked to the notification obligation under Article 52(1)'s first sentence.[458] It is crucial to note that the notification obligation is triggered not only when the compute threshold is actually met, but also – as supported by various compelling arguments – when the provider already knows that it will be met.[459]

121. Article 51(2)'s presumption is rebuttable.[460] This is underscored by Article 52(2), which sets out a procedure for providers to contest Article 51(1)(a)'s classification of GPAI models as presenting systemic risk together with notification.[461] This procedure arguably allows the provider to present arguments that, although the model meets or will meet the compute threshold, it does not have high-impact capabilities and therefore should not be classified as presenting systemic risk.[462]

122. To rebut the presumption of high-impact capabilities, the provider must prove that the model's capabilities do not meet the definition of high-impact capabilities under Article 3(64) – that is, that they do not 'match or exceed the capabilities recorded in the most advanced general-purpose AI

[453] Commission Guidelines (n 39) para 120.
[454] AI Act, art 52(1), first and second sentences; see commentary on Article 52, Section 2.1.2. in this work.
[455] Commission Guidelines (n 39) para 31; see Bernsteiner and Schmitt, 'Art. 52' (n 17) para 16; opposing view: Haar and Siglmüller, 'Art. 52' (n 24) para 8; further, see commentary on Article 52, Section 2.1.2. in this work.
[456] AI Act, art 51(2).
[457] See Section 2.1.1.1.
[458] See Commission Guidelines (n 39) paras 29, 31; Haar and Siglmüller, 'Art. 52' (n 24) para 6; see also commentary on Article 52, Section 2.1.1.2.1. in this work.
[459] Commission Guidelines (n 39) para 30 and Bernsteiner and Schmitt, 'Art. 52' (n 17) paras 13–14; opposing view: Haar and Siglmüller, 'Art. 52' (n 24) para 6; this is discussed in-depth in commentary on Article 52, Section 2.1.1.2.3. in this work.
[460] Commission Guidelines (n 39) para 33–36; Eric Hilgendorf and Johannes Härtlein, 'Art. 51 Einstufung von KI-Modellen mit allgemeinem Verwendungszweck als KI-Modelle mit allgemeinem Verwendungszweck mit systemischem Risiko' in Eric Hilgendorf and Johannes Härtlein (eds.), *KI-VO: Verordnung über künstliche Intelligenz* (Nomos 2025) para 2; Feiler, Forgó and Nebel, 'Article 51' (n 287) para 4; commentary on Article 52, Section 2.2.2.1. in this work. One may note that a provider of a GPAI model that meets the training compute threshold but exceptionally lacks high-impact capabilities may rebut this presumption only together with its notification pursuant to article 52(1)'s first sentence rather than choosing not to notify at all (see commentary on Article 52, Section 2.1.1.2.1. in this work).
[461] See Commission Guidelines (n 39) para 33–36; further, see Schneider and Schneider, 'Art. 51' (n 4) para 30. The procedure to contest classification, including the rebuttal of the article 51(2) presumption, and its corresponding requirements are set out in detail in commentary on Article 52, Section 2.2. in this work.
[462] Commission Guidelines (n 39) paras 33; see commentary on Article 52, Section 2.2.2.1. in this work.

models'.[463] To establish this, the provider can make use of appropriate technical tools and methodologies referred to under Article 51(1)(a),[464] with the Commission Guidelines specifically mentioning actual and forecasted benchmark results as potential grounds for rebuttal of the presumption.[465] Moreover, the provider may also argue that the fact that the model exceeds the threshold does not, in the particular circumstances, indicate high-impact capabilities.[466] This could be the case where the cumulative amount of training compute barely surpasses the threshold set by Article 51(2)[467] or where it is only met due to the inclusion of certain computational activities with a particularly low compute-to-capability ratio.

2.3. Article 51(3): Delegated acts

123. Article 51(3) provides that '[t]he Commission shall adopt delegated acts in accordance with Article 97 to amend the thresholds listed in paragraphs 1 and 2 of this Article, as well as to supplement benchmarks and indicators in light of evolving technological developments, such as algorithmic improvements or increased hardware efficiency, when necessary, for these thresholds to reflect the state of the art'.[468] This delegation of power to the Commission is based on Article 290(1) TFEU,[469] and the powers to 'amend' and to 'supplement' correspond to the two distinct categories of delegated powers laid down in this provision.[470] In general, a power to 'amend' a legislative act aims to 'authorise the Commission to modify or repeal non-essential elements' of an act, whereas a power to 'supplement' a legislative act aims to 'authorise the Commission to flesh out that act'.[471] Like other provisions of the AI Act empowering the Commission to adopt delegated acts, Article 51(3) aims to allow for necessary updates to the regulatory framework.[472]

2.3.1. Scope of the delegation of power

124. As laid out above, the AI Act refers to a multitude of assessment instruments in the context of classification, such as (technical) tools, methodologies, indicators, benchmarks, criteria, thresholds,

---

[463] See Commission Guidelines (n 39) paras 36–38. For a discussion of the definition of high-impact capabilities see forthcoming commentary on Article 3(64) in this work.

[464] See Section 2.1.1.2.

[465] Commission Guidelines (n 39) paras 34, 38; see Schneider and Schneider, 'Art. 51' (n 4) para 30 who argue that the annex XIII criteria may play a role in this context as well.

[466] Commission Guidelines (n 39) para 38.

[467] Commission Guidelines (n 39) para 38.

[468] See AI Act, recital 111, eighth sentence; recital 173, first sentence; and recital 179, seventh sentence. Schneider and Schneider, 'Art. 51' (n 4) para 40 express concern about the legal uncertainty that providers may face where delegated acts amend the relevant criteria at short notice.

[469] See Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C 326/47 ("TFEU") art 290(1): 'A legislative act may delegate to the Commission the power to adopt non-legislative acts of general application to supplement or amend certain non-essential elements of the legislative act.'; see also AI Act, recital 173, first sentence.

[470] See Case C-286/14 *European Parliament v European Commission* [2016] ECLI:EU:C:2016:183 para 40; Case C-617/24 Siegfried PharmaChemikalien Minden v Hauptzollamt Bielefeld [2025] ECLI:EU:C:2025:908 paras 24; Clara Saillant, 'Article 97 Exercise of the Delegation' in Ceyhun Necati Pehlivan, Nikolaus Forgó and Peggy Valcke (eds), *The EU Artificial Intelligence (AI) Act: A Commentary* (Wolters Kluwer 2024), 1339, s 3.2.

[471] See *European Parliament v European Commission* (n 470) paras 41–42; *Siegfried PharmaChemikalien Minden v Hauptzollamt Bielefeld* (n 470) para 30; further, see Non-Binding Criteria for the application of Articles 290 and 291 of the Treaty on the Functioning of the European Union [2019] OJ C 223/1 ("Non-Binding Criteria") s II.B. and C; Saillant (n 470) 1339, sec 3.2.

[472] AI Act, recital 173, first sentence.

approximations and evaluations, without defining them or clearly distinguishing between them – and potentially using some of them synonymously.[473] This inconsistent terminology creates uncertainty regarding the scope of Article 51(3)'s delegation of power, which specifically mentions thresholds, benchmarks and indicators.

125. As a result, some scholars have interpreted the scope of the delegation of power under Article 51(3) broadly, encompassing all relevant criteria under Article 51,[474] while other scholars see the scope of Article 51(3) more narrowly, applicable in particular to Article 51(2)'s training compute threshold.[475] The Commission Guidelines, staying close to Article 51(3)'s wording, set out that the provision empowers the Commission to 'adjust the thresholds set out in Article 51(1) and (2) AI Act' and 'to introduce additional benchmarks and indicators'.[476]

126. The subsequent sections examine the scope of Article 51(3)'s delegation of power with regard to (i) Article 51(2)'s training compute threshold;[477] (ii) Article 51(1)'s substantive criteria for classification;[478] (iii) indicators and benchmarks referred to under Article 51(1)(a);[479] and (iv) the criteria contained within Annex XIII.[480]

2.3.1.1. Training compute threshold under Article 51(2)

127. Based on the wording of Article 51(3) and the relevant recitals, it is clear that Article 51(3) empowers the Commission to amend Article 51(2)'s training compute threshold.[481] The examples of technological developments stated in Article 51(3) that could necessitate an update of this threshold – 'algorithmic improvements' and 'increased hardware efficiency' – suggest that the legislature primarily contemplated a lowering of the threshold. However, neither Article 51(3) nor the recitals exclude a raising of the threshold.[482]

2.3.1.2. Substantive criteria for classification under Article 51(1)

128. Some scholars argue that Article 51(3) also empowers the Commission to amend the substantive criteria for classification listed under Article 51(1).[483] Such a broad interpretation of Article 51(3) –

---

[473] See Section 2.1.1.2.; see also Haar and Siglmüller, 'Art. 51' (n 9) para 84.

[474] Bernsteiner and Schmitt, 'Art. 51' (n 6) para 40.

[475] Haar and Siglmüller, 'Art. 51' (n 9) para 82, leaving open whether article 51(3)'s delegation of power extends beyond article 51(2)'s training compute threshold to benchmarks and indicators contained in annex XIII or benchmarks referred to in article 66, point (g).

[476] Commission Guidelines (n 39) para 28.

[477] See Section 2.3.1.1.

[478] See Section 2.3.1.2.

[479] See Section 2.3.1.3.

[480] See Section 2.3.1.4.

[481] AI Act, art 51(3) and recital 111, seventh and eighth sentence. Haar and Siglmüller, 'Art. 51' (n 9) para 56; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 40; European Commission, 'General-Purpose AI Models in the AI Act – Questions & Answers' (n 4).

[482] See AI Act, recital 111, eighth sentence: '[Article 51(2)'s] threshold should be adjusted over time to reflect technological and industrial changes, such as algorithmic improvements or increased hardware efficiency, and should be supplemented with benchmarks and indicators for model capability.'

[483] In favour: Bond and Abbady, 'Art. 51' (n 138) 834, s 3.3 ('The Commission may also adopt delegated acts to amend the threshold of "high-impact capabilities" more generally [...].'); Carey (n 182) 10; Bernsteiner and Schmitt, 'Art. 51' (n 6) para 40; against: Haar and Siglmüller, 'Art. 51' (n 9) para 82; unclear: Schneider and Schneider, 'Art. 51' (n 4) 39; see also European Commission, 'General-Purpose AI Models in the AI Act – Questions & Answers' (n 4): 'For example, the value of the threshold [under Article 51(2)] itself could be adjusted and/or additional thresholds introduced.'

that requires interpreting these substantive criteria as 'thresholds listed in paragraphs 1 and 2' in the sense of Article 51(3) – could have profound implications, as it would grant the Commission the power to introduce new conditions for the classification of a GPAI model as presenting systemic risk besides a models' high-impact capabilities[484], or its capabilities or impact equivalent to high-impact capabilities[485]. This, in turn, could lead to the classification of new or different models as presenting systemic risk that may currently not be within scope of the classification conditions under Article 51(1).[486]

129. This broad interpretation does not exceed the wording of Article 51(3), as a substantive criterion such as the model's high-impact capabilities can be referred to as a qualitative threshold.[487] Indeed, Recital 111 can be read in such a way.[488] Moreover, this interpretation would explain why Article 51(3)'s power to amend thresholds does not refer only to the training compute threshold in Article 51's second paragraph but to its first paragraph as well – which otherwise remains difficult to explain.

130. In particular, it is not entirely convincing to interpret the power 'to amend thresholds listed in paragrap[h] 1' under Article 51(3) as referring to the assessment instruments for evaluating whether a model has high-impact capabilities mentioned in Article 51(1)(a).[489] Article 51(1)(a) does not mention thresholds but 'technical tools and methodologies, including indicators and benchmarks'.[490] Moreover, Article 51(3) refers to the power 'to amend thresholds listed in paragrap[h] 1' separate from the power 'to supplement benchmarks and indicators'.[491] Since the latter refers to assessment instruments mentioned in Article 51(1)(a),[492] interpreting the former as referring to them as well is not readily apparent. Additionally, if the legislature had intended the power 'to *amend* thresholds listed in paragrap[h] 1' (emphasis added) to refer to Article 51(1)(a)'s assessment instruments, it would be difficult to explain why it granted the power to 'amend' rather than to 'supplement'

[484] See AI Act, art 51(1)(a).

[485] See AI Act, art 51(1)(b.)

[486] From a practical perspective, the significance of different criteria for classification under Article 51(1) depends on the degree to which their application is shaped by additional thresholds, indicators and benchmarks.

[487] The notion of 'threshold' is not limited to quantitative thresholds under EU law (see, for example, Directive (EU) 2024/1203 of the European Parliament and of the Council of 11 April 2024 on the protection of the environment through criminal law and replacing Directives 2008/99/EC and 2009/123/EC [2024] OJ L 1203 recitals 9, 13 and 25 ('the qualitative and quantitative thresholds used to define environmental criminal offences') and Commission Delegated Regulation (EU) 2023/2772 of 31 July 2023 supplementing Directive 2013/34/EU of the European Parliament and of the Council as regards sustainability reporting standards [2023] OJ L 2772 para 42 ('quantitative and/or qualitative thresholds').

[488] Recital 111's eleventh sentence states that 'there should be a possibility for the Commission to take individual decisions designating a general-purpose AI model as a general-purpose AI model with systemic risk if it is found that such model has capabilities or an impact equivalent to those captured by the set threshold.' This sentence clearly mirrors the language of Article 51(1)(b), except for referring to 'those captured by the set threshold' instead of 'those set out in point (a)'. This implies that the 'set threshold' refers to the high-impact capabilities mentioned in article 51(1)(a) instead of the FLOPs threshold under article 51(2). The argument is nuanced, however, by the fact that the preceding sentence apparently refers to thresholds as assessment instruments (see AI Act, recital 111, tenth sentence: 'Thresholds, as well as tools and benchmarks for the assessment of high-impact capabilities should [...]'). As laid out above, the AI Act's terminology with regard to assessment instruments for systemic risk classification does not always appear entirely clear and consistent (see Section 2.1.1.2.).

[489] For a discussion of the assessment instruments referred to under article 51(1)(a), see Section 2.1.1.2.; for an analysis of the Commission's power to 'supplement benchmarks and indicators', see Section 2.3.1.3.

[490] However, the AI Act does not clearly distinguish between its denominations for different assessment instruments (Haar and Siglmüller, 'Art. 51' (n 9) para 84; see also Section 2.1.1.3.).

[491] For the latter see Section 2.3.1.3.

[492] See Section 2.3.1.3.

thresholds in this context.[493] The introduction of new thresholds as assessment instruments for evaluating a model's high-impact would constitute – like the introduction of benchmarks and indicators – a matter of fleshing out Article 51(1)(a)'s provision on assessment instruments rather than modifying or repealing it.[494] Nor is it entirely convincing to interpret the power 'to amend the thresholds listed in paragrap[h] 1' as referring to the business user threshold contained in point (f) of Annex XIII, since Article 51(1)(b) contains only a reference to this annex while the threshold itself is not contained in Article 51(1).[495] In light of these considerations, an interpretation of Article 51(3) as empowering the Commission to amend the substantive criteria for classification listed under Article 51(1) could ensure that the provision for the Commission's power 'to amend thresholds listed in paragrap[h] 1' retains its effectiveness.[496]

131. However, it is questionable to what extent the power to determine the substantive requirements for classification of a GPAI models with systemic risk can be delegated. Article 290(1) TFEU requires that '[t]he essential elements of an area shall be reserved for the legislative act and accordingly shall not be the subject of a delegation of power'. Whether and to what extent the substantive requirements under Article 51(1) constitute 'essential elements' within the meaning of Article 290(1) TFEU appears highly uncertain.[497]

132. Further, a purposive argument may be advanced against this broad interpretation of Article 51(3) as allowing for the amendment of Article 51(1)(a)'s high-impact capabilities requirement. The legislature has already shaped the concept of high-impact capabilities in a way that is responsive to evolving technological developments by defining it in relation to the 'capabilities recorded in the most advanced general-purpose AI models'.[498] Consequently, it remains uncertain whether an amendment of the substantive criteria under Article 51(1) can be 'necessary [...] in light of evolving

---

[493] In general, a power to 'amend' a legislative act aims to 'authorise the Commission to modify or repeal non-essential elements' of an act, whereas a power to 'supplement' a legislative act aims to 'authorise the Commission to flesh out that act' (see *European Parliament v European Commission* (n 470) paras 41–42; *Siegfried PharmaChemikalien Minden v Hauptzollamt Bielefeld* (n 470) para 30; Non-Binding Criteria (n 471) ss II.B. and C; Saillant (n 470) 1339, s 3.2; see also Section 2.3.1.).

[494] Compare *European Parliament v European Commission* (n 470) paras 41–42; *Siegfried PharmaChemikalien Minden v Hauptzollamt Bielefeld* (n 470) para 30; Non-Binding Criteria (n 471) ss II.B. and C.

[495] See Section 2.3.1.4.

[496] See, for example, *RW v Österreichische Post AG* (n 159) para 29 and the case law cited therein: '[W]here a provision of EU law is open to several interpretations, preference must be given to that interpretation which ensures that the provision retains its effectiveness.'

[497] According to case law, 'ascertaining which elements of a matter must be categorised as "essential" is not for the assessment of the EU legislature alone, but must be based on objective factors amenable to judicial review. Account must be taken of the characteristics and particular features of the field concerned [...].' (Case C-696/15 P *Czech Republic v Commission* [2017] ECLI:EU:C:2017:595, para 77 and the case law cited) The Court further held that '[a]n element is essential within the meaning of the second sentence of the second subparagraph of Article 290(1) TFEU in particular if, in order to be adopted, it requires political choices falling within the responsibilities of the EU legislature, in that it requires the conflicting interests at issue to be weighed up on the basis of a number of assessments, or if it means that the fundamental rights of the persons concerned may be interfered with to such an extent that the involvement of the EU legislature is required [...].' (*Czech Republic v European Commission* para 78). This standard appears rather flexible, and it is unclear what the result of its application in the present context would be. An argument can be made that the substantive requirements for classification constitute an essential element of GPAI model regulation under the AI Act as they determine the applicability of the specific rules for GPAI models with systemic risk (see Section 2.1.4.). However, one could also contend that the requirements for classification under Article 51(1)(a) may not be essential within the meaning of Article 290(1) TFEU because they do not definitively determine a model's classification as the provider may present arguments against classification pursuant to article 52(2) where the condition under article 51(1)(a) is met (see commentary on Article 52, Section 2.2. in this work).

[498] AI Act, art 3(64).

technological developments, such as algorithmic improvements or increased hardware efficiency' in the sense of Article 51(3).[499]

133. In case of an amendment of the substantive criteria for classification under Article 51(1)(b), this would, in principle, necessitate an amendment of Annex XIII's introductory wording as it mirrors Article 51(1)(b)'s language.[500] Interestingly, Article 52(4)'s second subparagraph by its wording only empowers the Commission to specify and update the criteria but not the chapeau of Annex XIII. Moreover, it is uncertain whether Article 51(3) itself allows for an amendment of Annex XIII.[501]

2.3.1.3. Indicators and benchmarks under Article 51(1)(a)

134. Article 51(1)(a) mentions not only the substantive criterion of a model's high-impact capabilities but also 'indicators and benchmarks' for their evaluation. Article 51(3) empowers the Commission to introduce and specify these assessment instruments via a delegated act.[502]

2.3.1.4. Criteria contained within Annex XIII

135. Whether and to what extent the delegation of power under Article 51(3) encompasses thresholds, benchmarks and indicators contained in Annex XIII, is unclear.[503] This question matters: although Article 52(4)'s second subparagraph specifically empowers the Commission to amend Annex XIII by specifying and updating the criteria, that only establishes the Commission's *power* to adopt delegated acts, whereas Article 51(3) arguably obliges the Commission to make use of the corresponding delegation of power under certain circumstances.[504] Such an *obligation* could be particularly relevant with regard to point (f) of Annex XIII. This provision contains a quantitative threshold as it provides that a model's high impact on the internal market due to its reach shall be presumed when it has been made available to at least 10,000 registered business users established in the Union.[505] Thresholds and benchmarks are further expressly mentioned in points (d) and (e) of Annex XIII.[506]

---

[499] It is conceivable that the evaluation of whether a GPAI model has high-impact capabilities may become more difficult due to technological developments or training compute becoming a less relevant indicator for high-impact capabilities. In such cases, an amendment of article 51(1) could potentially be seen as necessary for ensuring that article 51(1) provides for practically operable conditions for classification. However, scenarios in which it could be necessary to amend the substantive criteria in article 51(1) appear less tangible than scenarios in which it is necessary to adjust the quantitative threshold of floating-point operations under article 51(2).

[500] See AI Act, annex XIII: 'For the purpose of determining that a general-purpose AI model has capabilities or an impact equivalent to those set out in Article 51(1), point (a), [...]'.

[501] See Section 2.3.1.4.

[502] Hofmann-Coombe (n 6) para 53; Martini (n 2) para 195; see recital 111's eighth sentence, which states: 'This threshold [of floating point operations] should be adjusted over time to reflect technological and industrial changes, such as algorithmic improvements or increased hardware efficiency, and should be supplemented with benchmarks and indicators for model capability.'

[503] In favour: Bernsteiner and Schmitt, 'Art. 51' (n 6) paras 40–41; Hilgendorf and Härtlein, 'Art. 51' (n 460) para 6; see also Haar and Siglmüller, 'Art. 51' (n 9) para 83 who doubt whether article 51(3) refers to annex XIII in light of the delegation of power under Article 52(4)'s second subparagraph.

[504] See Section 2.3.2.

[505] For this presumption's role in classification under Article 51(1)(b), see Section 2.1.2.1.4.

[506] See AI Act, annex XIII: 'For the purpose of determining that a general-purpose AI model has capabilities or an impact equivalent to those set out in Article 51(1), point (a), the Commission shall take into account the following criteria: [...] (d) the input and output modalities of the model, such as text to text (large language models), text to image, multi-modality, and *the state of the art thresholds for determining high-impact capabilities for each modality*, and the specific type of inputs and outputs (e.g. biological sequences); (e) the *benchmarks* and evaluations *of*

136. However, extending the delegation of power under Article 51(3) to *thresholds* contained in Annex XIII is not close at hand since this delegation provision specifically refers to thresholds listed in Article 51's first and second paragraph.[507] With regard to *benchmarks* and *indicators* contained in Annex XIII, it is difficult to envisage that their supplementation would ever be necessary in the sense of Article 51(3), particularly if one views the list of criteria as non-exhaustive.[508] Moreover, Annex XIII refers to its content as 'criteria' rather than as thresholds, benchmarks or indicators.[509] Since Article 51(3) – unlike Article 52(4)'s second subparagraph[510] – neither uses the term 'criteria' nor expressly references Annex XIII, this provides a further argument against the delegation of power under Article 51(3) encompassing criteria contained in Annex XIII.

2.3.2. Obligation to adopt delegated acts

137. The wording of Article 51(3) ('the Commission *shall* adopt delegated acts') suggests an obligation of the Commission to make use of the powers delegated to it under certain instances.[511] Such an obligation to adopt delegated acts, recognised under EU law more generally,[512] is supported by comparison to the wording of other delegation provisions in Chapter V. Under Article 52(4)'s first subparagraph and Article 53(5) and (6), the Commission '*is empowered* to adopt delegated acts' (emphasis added) to update Annex XIII and Annexes XI and XII, respectively. In general, the distinction between *shall* and *is empowered* is not merely of stylistic nature because *shall* generally imposes a binding obligation, while *being empowered* signals discretion that the Commission may exercise.[513]

---

capabilities of the model, including considering the number of tasks without additional training, adaptability to learn new, distinct tasks, its level of autonomy and scalability, the tools it has access to; [...]' (emphasis added).

[507] The contrary view would need to rely on the argument that annex XIII is expressly mentioned in article 51(1)(b), and thus a threshold listed in annex XIII is indirectly listed in article 51(1) as well. A trilogue preparation document indeed suggests that article 51(3)'s reference of thresholds was intended to cover the business user threshold contained in point (f) of annex XIII as well at an earlier drafting stage (see Council of the European Union, 'AI Act – Preparation for the trilogue' (Note from the Presidency to the Permanent Representatives Committee, 16097/23, 28 November 2023) (Interinstitutional File 2021/0106(COD)) 8).

[508] See Section 2.1.2.1.2.2.

[509] This is regardless of the fact that these criteria contain thresholds, benchmarks and indicators (see, for example, AI Act, annex XIII, points (d)–(f)).

[510] See AI Act, art 52(4), first subparagraph: 'The Commission is empowered to adopt delegated acts in accordance with Article 97 in order to amend Annex XIII by specifying and updating the *criteria set out in that Annex*.' (emphasis added)

[511] See Bernsteiner and Schmitt, 'Art. 51' (n 6) paras 40–41 who derive from article 51(3) the Commission's obligation for market observation and implementation of a continuous consultation procedure in order to be able to assess when it is necessary to adapt the criteria of article 51 and annex XIII; the AI Act further provides that the Commission 'shall adopt delegated acts' under articles 6(7), 43(6). Also see (seemingly opposing view) Haar and Siglmüller, 'Art. 51' (n 9) para 78 who describe the Commission as 'empowered' ('ermächtigt') to adopt such acts.

[512] See Case T-55/24 *Meta Platforms Ireland Ltd v European Commission* [2025] ECLI:EU:T:2025:842, paras 43–44 where the General Court found that article 43(4) DSA, which employs a comparable wording ('The Commission shall adopt delegated acts [...]'), provided for an obligation of the Commission to adopt a delegated act. In Case C-137/21 *European Parliament v European Commission* [2023] ECLI:EU:C:2023:625, para 56–64, the Court of Justice found that a provision's wording stating that the Commission 'shall adopt' a delegated act implies that 'the Commission is required to adopt such an act where the conditions required for its adoption are satisfied' (para 57) but ruled out the existence of such an obligation in the specific case of article 7(1), point (f) of Regulation (EU) 2018/1806 of the European Parliament and of the Council of 14 November 2018 listing the third countries whose nationals must be in possession of visas when crossing the external borders and those whose nationals are exempt from that requirement [2018] OJ L 303/39 based on the provision's context and the objectives of the legislation.

[513] See European Commission, 'English Style Guide: A Handbook for Authors and Translators in the European Commission' (2025) para 10.27: 'To impose an obligation or a requirement, EU legislation uses shall.'

138. The AI Act's recitals support this distinction in the present case, which state with regard to Article 51(3) that the 'threshold [of floating point operations under Article 51(2)] *should be adjusted* over time to reflect technological and industrial changes, such as algorithmic improvements or increased hardware efficiency, and *should be supplemented* with benchmarks and indicators for model capability' (emphasis added).[514] This interpretation is in line with the Commission Guidelines which state with respect to the technical tools and methodologies mentioned under Article 51(1)(a) that '[t]hese tools and methodologies *are to be further specified* by the Commission through adoption of delegated acts [...]' (emphasis added).[515]

139. The Commission will likely enjoy considerable discretion in determining whether this obligation is triggered. The legislature has not tied the duty to exercise the delegated power to specific timeframes but made it contingent upon whether it is 'necessary, for these thresholds to reflect the state of the art'. Nonetheless, there are indications that the legislature intended only a limited number of providers of GPAI models with systemic risk to be classified at a time.[516] Article 265 TFEU provides for an action for failure to act before the Court of Justice of the European Union to have the infringement established.[517]

140. It is unlikely that a provider could challenge the classification of its model as presenting systemic risk on the grounds that the Commission breached its obligation to amend the threshold. Such reasoning would effectively make the threshold's applicability contingent upon the Commission's fulfilment of its obligation under Article 51(3), while the legislature did not expressly provide for Article 51(2)'s threshold's conditional applicability.[518] Nor is there any indication that such conditional applicability of Article 51(2)'s threshold was intended. The example scenarios mentioned in Article 51(3) – 'algorithmic improvements or increased hardware efficiency' – suggest that the legislature primarily envisaged a lowering of the training compute threshold.[519] In such scenarios, the relevant concern would not be whether models were wrongly classified under an outdated threshold, but rather whether models escaped classification that nevertheless present systemic risks despite having been trained with lower compute, for example because of algorithmic improvements. Moreover, meeting Article 51(2)'s threshold merely gives rise to a presumption of high-impact capabilities, which its provider can rebut by contesting the model's classification under Article 52(2)–(3).[520] This contestation procedure allows a provider to prevent its model's classification based on a potentially outdated threshold, where the model does not present systemic

---

[514] AI Act, recital 111, eighth sentence. This contrasts with the language of recital 101's fifth sentence, which states with regard to article 53(6): 'The Commission should be empowered to amend those annexes by means of delegated acts in light of evolving technological developments'. See also Bernsteiner and Schmitt, 'Art. 51' (n 6) para 41.

[515] Commission Guidelines (n 39) para 27.

[516] See para 3.

[517] See Case T-521/14 *Kingdom of Sweden v European Commission* [2015] ECLI:EU:T:2015:976; further, see *European Parliament v European Commission* (n 512) where the Court dismissed the Parliament's action for failure to act based on the Commission's alleged infringement of the Treaties by failing to adopt, pursuant to point (f) of the first paragraph of article 7 of Regulation 2018/1806, a delegated act temporarily suspending the exemption from the visa requirement for nationals of the United States of America as unfounded.

[518] Examples where the legislature made the applicability of EU law provisions contingent upon the adoption of a delegated act include Regulation (EU) No 909/2014 (n 116), art 76(5) ('The settlement discipline measures referred to in Article 7(1) to (13) and the amendment laid down in Article 72 shall apply from the date of entry into force of the delegated act adopted by the Commission pursuant to Article 7(15).'); Regulation (EU) 2015/2365 (n 116), art 33(2)(a) ('Article 4(1) [...] shall apply: (i) 12 months after the date of entry into force of the delegated act adopted by the Commission pursuant to Article 4(9) [...]').

[519] See also Section 2.3.1.1.

[520] For the effects of article 51(2)'s presumption, see Section 2.2.2.; for the procedure to contest classification under article 52(2) and (3), see commentary on Article 52, Section 2.2. in this work.

risk.[521] In particular, in the context of this contestation procedure, the Commission takes into account how indicative the model's training compute is for assessing whether it has high-impact capabilities.[522]

2.3.3. Conditions for the adoption of delegated acts

141. The power to adopt delegated acts under Article 51(3) is conferred on the Commission subject to the conditions laid down in Article 97.[523] The delegation runs for five years from 1 August 2024, with tacit extension for periods of an identical duration absent opposition by the European Parliament and the Council no later than three months before the end of each period.[524] Both institutions retain the power to revoke the delegation at any time.[525] Before adopting any delegated act, the Commission must consult Member State experts designated by the Member States in line with the principles established in the Interinstitutional Agreement on Better Law-Making of 13 April 2016.[526] Once adopted, the delegated act must be notified to the European Parliament and the Council,[527] which then have three months (extendable to six months) to raise objections that would prevent the act from taking effect.[528]

2.4. Annex XIII

2.4.1. Overview

142. Annex XIII contains a heterogeneous list of around eleven[529] criteria with relevance for systemic risk classification of GPAI models.[530] These criteria are particularly relevant for Commission designation decisions under Article 51(1)(b) in conjunction with Article 52(4)'s first subparagraph, where the Commission must take these criteria into account.[531] However, they may also play a role in other instances connected to the classification of GPAI models as presenting systemic risk, such as

[521] See commentary on Article 52, Section 2.2.2.1. in this work

[522] See Commission Guidelines (n 39) para 38: 'In its assessment of whether the model is amongst the most advanced models at the time of notification, the Commission will take into account the extent to which the cumulative training compute of the model is indicative of the model being amongst these models.'

[523] AI Act, art 51(3) in conjunction with AI Act, art 97(1).

[524] AI Act, art 97(2); see Christina Brandt-Steinke, 'Art. 97 Ausübung der Befugnisübertragung' in Jens Schefzig and Robert Kilian (eds), *Beck'scher Online-Kommentar KI-Recht* (4th edn, C H Beck 2025) paras 19–20; Michael Kolain, 'Art. 97 Ausübung der Befugnisübertragung' in Mario Martini and Christiane Wendehorst (eds), *KI-VO: Verordnung über Künstliche Intelligenz: Kommentar* (2nd edn, C H Beck 2026) para 20.

[525] AI Act, art 97(3); see Brandt-Steinke (n 524) paras 22–23; Kolain, Art. 97 (n 524) paras 22–26.

[526] AI Act, art 97(4); see Brandt-Steinke (n 524) paras 27–30; Kolain, Art. 97 (n 524) paras 27–28; further, see AI Act, recital 111, ninth sentence ('[T]he AI Office should engage with the scientific community, industry, civil society and other experts [when adjusting the threshold of floating point operations].') and AI Act, recital 173, second sentence ('It is of particular importance that the Commission carry out appropriate consultations during its preparatory work, including at expert level, and that those consultations be conducted in accordance with the principles laid down in the Interinstitutional Agreement of 13 April 2016 on Better Law-Making').

[527] AI Act, art 97(5); see Brandt-Steinke (n 524) para 31; Kolain, Art. 97 (n 524) para 29.

[528] AI Act, art 97(6); see Brandt-Steinke (n 524) para 33; Kolain, Art. 97 (n 524) para 30.

[529] The exact number depends on the counting method, for example, whether benchmarks and evaluations of capabilities of the model under point (e) of annex XIII are counted as one or two distinct criteria.

[530] See Commission Guidelines (n 39) para 5.

[531] See Bond and Abbady, 'Art. 51' (n 138) 833–834, s 3.2.

reassessment of a designation under Article 52(5) or the assessment of whether a GPAI model has high-impact capabilities under Article 51(1)(a).[532]

143. The criteria, which are grouped in seven points and already touched upon above,[533] range from rather specific, technical and quantifiable (e.g. the number of model parameters[534]) to rather unspecific (e.g. the evaluations of capabilities of the model[535]), non-technical (e.g. the number of registered end users[536]) and not easily quantifiable (e.g. the quality of the data set[537]). Some criteria are more related to a model's capabilities (e.g. the evaluations of the model's capabilities[538]), whereas others are more related to its impact (e.g. the number of registered end users[539]).[540] In some instances, Annex XIII only lists the criterion itself (e.g. the number of model parameters[541]), while in others it provides illustration through examples (e.g. biological sequences as a specific type of model input[542]), indications of measurement units (e.g. FLOPs for amount of computation used for training[543]), and measurement methods (e.g. measuring the size of the data set through tokens[544]). Regarding the model's high impact on the internal market due to its reach, point (f) of Annex XIII notably contains a presumption based on the number of registered business users.[545]

144. Article 52(4)'s second subparagraph allows the Commission to 'specif[y] and updat[e]' the Annex XIII criteria.[546] In this respect, the more compelling arguments appear to support the view that the Commission is also empowered to add new criteria to the list.[547] Beyond this, it appears rather doubtful whether Article 51(3)'s delegation of power extends to Annex XIII as well.[548] This could be relevant insofar as the Commission might, under certain conditions, be obliged to adapt the threshold of 10,000 business users under point (f) of Annex XIII to technological developments.[549]

2.4.2. Relevance for classification

145. Annex XIII has overarching significance for the classification rules in Section 1 of Chapter V,[550] which expressly mentions it four times. Article 51(1)(b) serves as Annex XIII's primary anchor

[532] See Section 2.4.2.
[533] See, in particular, Section 2.1.2.1.2.3. for points (a) to (e) and Section 2.1.2.1.4. for points (f) and (g) of annex XIII.
[534] AI Act, annex XIII, point (a).
[535] AI Act, annex XIII, point (e).
[536] AI Act, annex XIII, point (g).
[537] AI Act, annex XIII, point (b).
[538] AI Act, annex XIII, point (e).
[539] AI Act, annex XIII, point (g).
[540] See Bond and Abbady, 'Art. 51' (n 138) 832–833, s 3.2; Hacker, Kasirzadeh and Edwards (n 61) 15.
[541] AI Act, annex XIII, point (a); see also AI Act, annex XIII, point (g).
[542] AI Act, annex XIII, point (d); see also AI Act, annex XIII, point (e).
[543] AI Act, annex XIII, point (c); see also AI Act, annex XIII, point (b).
[544] AI Act, annex XIII, point (b); see also AI Act, annex XIII, point (c).
[545] AI Act, annex XIII, point (g); for the role of this presumption for classification under Article 51(1)(b), see Section 2.1.2.1.4.
[546] See commentary on Article 52, Section 2.3.2. in this work.
[547] See commentary on Article 52, Section 2.3.2. in this work.
[548] See Section 2.3.1.4.
[549] See Section 2.3.1.4.
[550] See Oskar J. Gstrein, Noman Haleem and Andrej Zwitter, 'General-Purpose AI Regulation and the European Union AI Act' (2024) 13 Internet Policy Review s 3 ('The criteria for this classification [...] will need to be interpreted and updated by regulators along the 7 criteria provided in Annex XIII.'); see also Commission Guidelines (n 39), para 5, which mention annex XIII without a reference to a specific provision under Section 1 of Chapter V ('The

point,[551] as confirmed both by Annex XIII's title referring to Article 51 and by the fact that Article 51(1)(b) is the only provision in Article 51 that expressly references the Annex. Annex XIII's wording reinforces this connection by stating that it contains a list of criteria that the Commission shall take into account '[f]or the purpose of determining that a general-purpose AI model has capabilities or an impact equivalent to those set out in Article 51(1), point (a)' – language that mirrors Article 51(1)(b). The express reference to Article 51(1)(a) in Annex XIII should therefore not be misunderstood as indicating that Annex XIII relates directly to that provision[552] – rather, the reference reflects the fact that Article 51(1)(a) is itself expressly mentioned in Article 51(1)(b).

146. Annex XIII's relevance to classification is further reflected in Article 52 as well. Article 52(4)'s first subparagraph establishes the Commission's power to designate GPAI models as presenting systemic risk 'on the basis of criteria set out in Annex XIII',[553] while Article 52(5)'s first sentence clarifies that reassessment of such designations must likewise be made on the basis of Annex XIII's criteria.[554] The reference to Annex XIII in Article 52(4)'s first subparagraph provides additional support for the view taken here that the designation decision under that provision coincides with the Commission decision referred to in Article 51(1)(b).[555]

147. While Annex XIII is not expressly mentioned in the provisions relating to the classification based on a model's high-impact capabilities, its criteria may still be considered when deciding whether a model satisfies Article 51(1)(a) or whether, due to its specific characteristics, it does not present systemic risks under Article 52(2) and (3).[556] Several criteria in Annex XIII appear to serve as indicators of a model's (high-impact) capabilities rather than its impact. Examples include the amount of computation used for the model's training – mentioned not only in Article 51(2) but also in Annex XIII, point (c) – and the benchmarks and evaluations of model capabilities referred to in point (e) of the Annex.

Commission may also designate general-purpose AI models as general-purpose AI models with systemic risk based on the criteria in Annex XIII AI Act.')

[551] For the role of annex XIII's criteria for classification under article 51(1)(b), see Section 2.1.2.1.2.

[552] However, this is suggested by Feiler, Forgó and Nebel, 'Article 51' (n 287) para 6 ('The evaluation referred to in para. 1(a) [...] must take into account the criteria of Annex XIII (Annex XIII sentence 1)'); see also Bernsteiner and Schmitt 'Art. 51' (n 6) para 25 ('Das bedeutet natürlich nicht, dass der Anbieter des KI-Modells nicht ebenfalls die in Anh. XIII KI-VO festgelegten Kriterien bei seiner eigenen Prüfung (mit-)berücksichtigen kann. Dies zumal der Anh. in seinem ersten Satz auf Art. 51 Abs. 1 lit. a verweist.'); unclear in that respect: Carey (n 182) 8 ('Annex XIII provides further criteria for the determination of systemic risk due to high-impact capabilities, as well as some guidance on the "indicators and benchmarks."').

[553] See commentary on Article 52, Section 2.3.1. in this work.

[554] See commentary on Article 52, Section 2.4. in this work.

[555] See commentary on Article 52, Section 2.1.3.1. and Section 2.3.1. in this work.

[556] Carey (n 182) 8 ('Annex XIII provides further criteria for the determination of systemic risk due to high-impact capabilities, as well as some guidance on the "indicators and benchmarks.'); Bond and Abbady, 'Art. 51' (n 138) 835, s 4.1; with regard to article 51(1)(a) see also Bernsteiner and Schmitt, 'Art. 52' (n 17) para 16 and Haar and Siglmüller, 'Art. 51' (n 9) para 32.