European Commission

Joint Research Centre

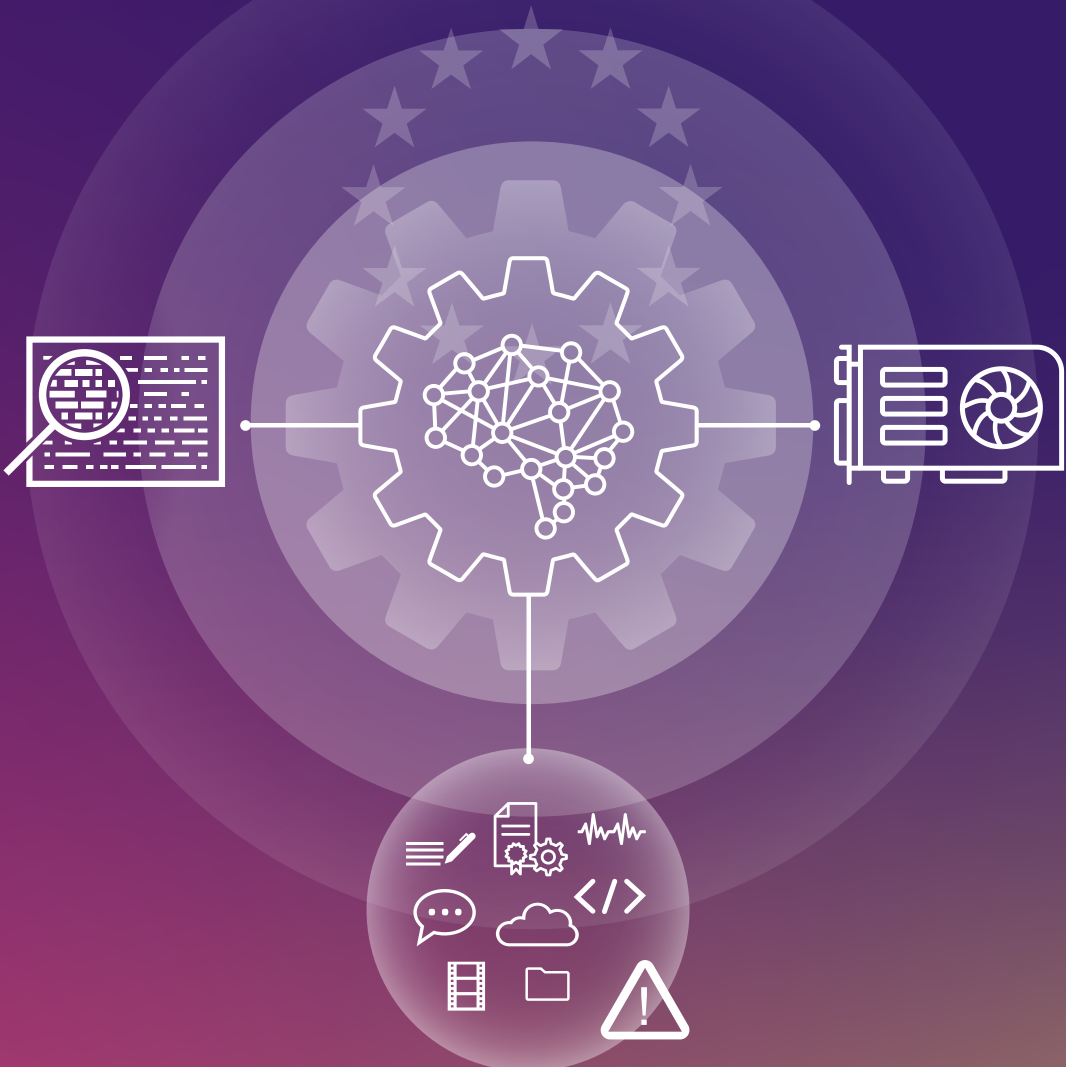
General-Purpose AI Model Reach as a Criterion for Systemic Risk

Collection of External Scientific Studies on General-Purpose AI Models under the EU AI Act

Röttger, P., Hovy, D., Nozza, D., Vanschoren, J.

Fernández Llorca, D., Noroozian, A., Gómez, E. (editors)

2025



Contact Information
Name: David Fernández Llorca
Address: European Commission, Joint Research Centre (JRC) Edificio Expo, c/Inca Garcilaso, 3, 41092 Seville - Spain
Email: david.fernandez-llorca@ec.europa.eu

The Joint Research Centre: EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC143260

PDF ISBN 978-92-68-31429-6 doi:10.2760/1853792 KJ-01-25-457-EN-N

Luxembourg: Publications Office of the European Union, 2025

How to cite this report: Röttger, P., Hovy, D., Nozza, D., Vanschoren, J., *General-Purpose AI Model Reach as a Criterion for Systemic Risk* Fernández Llorca, D., Noroozian, A., Gómez, E. (editors), Publications Office of the European Union, Luxembourg, 2025, https://data.europa.eu/doi/10.2760/1853792, JRC143260.

Contents

Abstract 2

Acknowledgements 3

Note from the Editors 4

Executive summary 6

1 Introduction 7

1.1 Legal Context 7

1.2 Risks Created by Reach 7

1.3 Risks from Reach in the Code of Practice 9

2 Metrics 11

2.1 Definition of Reach 11

2.2 Access Methods 11

2.3 Reach Metrics - User Interface Access 12

2.4 Reach Metrics - API Access 13

2.5 Measuring Reach using Public Information 14

2.6 Measuring Reach using Surveys 15

3 Concrete Proposal for the Design of a Reach Criterion 16

3.1 Overall Reporting Thresholds for Model Providers 16

3.2 Reach Metrics to be Reported 17

3.3 Calibrating the Reach Criterion 18

4 Conclusions 19

References 20

List of abbreviations and definitions 21

Abstract

Under the EU AI Act, systemic risks of general-purpose AI (GPAI) models are assumed to increase with model "reach". This report proposes a method for operationalising, and measuring "reach" so that it may be used, where appropriate, as one criterion amongst others for determining whether a GPAI model should be classified as a GPAI model with systemic risk. This report considers specific systemic risks created by model reach, such as risks from model biases, i.e. systemic risk which become relevant when models are widely used, regardless of whether their capabilities are at the technological frontier. In this report, reach is defined as the number of people who directly interact with a given GPAI model. Potential metrics for measuring reach are identified, including access via user interfaces and APIs. A concrete proposal for the design of a reach trigger is presented, which includes overall reporting thresholds for model providers, reach metrics to be reported, and calibration of the reach trigger. The proposal aims to balance comprehensive coverage with minimal reporting burden, and to provide a framework for incorporating reach considerations in the determination of whether a GPAI model presents systemic risks, in complement to other considerations related to capabilities, safety benchmarks and compute discussed in other reports of this collection.

Acknowledgements

The authors would like to thank the members of the systemic risk working group of the expert pool, as well as members of the Joint Research Centre and the AI Office, for their discussions and valuable feedback.

The editors would like to thank the JRC colleagues who have helped us develop the work resulting in this collection, the AI Office for its inputs, as well as those who have kindly agreed to review the drafts and the final external study reports.

Authors
Röttger, Paul
Hovy, Dirk
Nozza, Debora
Vanschoren, Joaquin

Editors
Fernández Llorca, David
Noroozian, Arman
Gómez, Emilia

Note from the Editors

The EU AI Act entered into force on 1 August 2024, with the aim of promoting innovation in and uptake of AI in the Union, while ensuring a high level of protection of health, safety and fundamental rights, including democracy and the rule of law. Chapter V of the AI Act outlines obligations for the providers of general-purpose AI (GPAI) models, which are AI models *"trained with a large amount of data using self-supervision at scale, that [display] significant generality and [are] capable of competently performing a wide range of distinct tasks [...] and that can be integrated into a variety of downstream systems or applications"*. Moreover, the chapter specifies additional obligations for the most advanced GPAI models, those that pose systemic risks, which are classified as such according to criteria established in Article 51 and Annex XIII. From 2 August 2025, the obligations for providers of GPAI models and GPAI models with systemic risk enter into application.

The European Commission's Joint Research Centre (JRC) has been providing scientific support throughout the legislative process of the AI Act since 2020. After the Council and Parliament reached a final agreement in December 2023, the JRC initiated an internal study, which included two external experts, focusing on the technical aspects of Chapter V that likely required further clarification. This study generated an internal report titled "General Purpose AI Models under the AI Act" (Hernández-Orallo et al., 2024), which provided preliminary insights into compute, generality, capabilities, and systemic risks. One of the clear conclusions of this preliminary study was that further scientific work was necessary.

Between September 2024 and June 2025, the JRC setup and managed, in close collaboration with the EU AI Office, a pool of 15 external experts with diverse expertise and backgrounds. This expert pool produced further technical scientific input on key aspects of Chapter V, through the development of methodologies for categorising AI models as GPAI models and for classifying GPAI models as GPAI models with systemic risk, to inform implementation of the EU AI Act. The experts also provided input on the recently published Commission guidelines on the scope of obligations for providers of GPAI models (European Commission, 2025), as part of the public multi-stakeholder consultation. The primary outcome of this expert pool is this **Collection of External Scientific Studies on General-Purpose AI Models under the EU AI Act**, which comprises a total of six external scientific study reports.

Although more documents may be added to the collection in the future, as of the writing of this editorial, the titles of the external reports included in the collection are:

— Training Compute Thresholds - Key Considerations for the EU AI Act

— A Framework for General-Purpose AI Model Categorisation

— A Framework to Categorise Modified General-Purpose AI Models as New Models Based on Behavioural Changes

— A Proposal to Identify High Impact Capabilities of General-Purpose AI Models

— The Role of AI Safety Benchmarks in Evaluating Systemic Risks in General-Purpose AI Model

— General-Purpose AI Model Reach as Criterion for Systemic Risk

The overall objective of this collection is twofold. On the one hand, it aims to contribute to broadening the understanding and discussion of the technical and scientific issues related to GPAI models and the identification of systemic risks. On the other hand, it seeks to provide a solid scientific basis for informing the implementation of Chapter V of the EU AI Act, which has recently entered into application. It is clear that we are dealing with complex issues, where a clear scientific consensus has yet to be established, and which require a certain degree of flexibility. Nevertheless, this is part of the necessary effort to promote innovation and the uptake of AI, while ensuring protection for human health, safety, and fundamental rights in Europe.

These external scientific studies cover aspects regarding the presumption of having high impact capabilities based on cumulative amount of computation used for training (Article 51(2)), notification conditions (Article 52), the definition of a GPAI model (Article 3(63)), and considerations for GPAI models being classified as GPAI models with systemic risk based on capability benchmarks, safety benchmarks and reach (Article 51(1) and Annex XIII).

These studies reflect the outcome of the scientific and technical analysis of a series of external experts to the Commission. In some cases, they present a state-of-the-art review, while in others, they propose methodologies based on solid scientific evidence, while acknowledging significant uncertainty, as many of these problems still lack a widely accepted solution. The content, analysis, recommendations, and suggestions should then not be interpreted in any way as the position of the Commission, nor of the JRC editors in particular, but rather as the opinion of the authors.

Executive summary

In this report, our goal is to address the following points:

— How might the "reach" of a general-purpose AI (GPAI) model be taken into account when determining whether a GPAI model present systemic risks? In particular, what are the best candidates for operationalising and measuring reach, and why?

— Any method for measuring reach should satisfy the property that it can reasonably easily be checked by model providers.

— What are the assumptions and uncertainties underlying the proposed approach?.

In this report, we argue that taking into account reach in determining whether a GPAI model is a GPAI model with systemic risk (hereafter "GPAISR model") is both legally feasible under the EU AI Act (European Union, 2024) and technically plausible. Doing so would address specific systemic risks created by reach. Under our threat model, these are primarily risks from model biases, which become relevant only when models are widely used, regardless of whether they are at the technological frontier.

The report is split into three main sections:

— Section 1: Motivation for taking into account risk when determining if a GPAI model present systemic risks. This section examines the legal context in the EU AI Act for why reach might be considered as a factor for systemic risks, describes risks specifically created by reach, and evaluates how these risks align with the third draft of the GPAI Code of Practice (Independent Experts, 2025)[1].

— Section 2: Metrics for measuring the reach of GPAI models. This section analyses potential metrics for measuring model reach, focusing on direct user interactions through both user interfaces (UIs) and application programming interfaces (APIs). The Section evaluates various measurement approaches, their advantages, and limitations.

— Section 3: Concrete proposal for how reach could be taken into account in determining whether a GPAI model should be a GPAI model with systemic risk. This section presents a specific implementation framework where model providers above certain thresholds would report monthly user counts to the enforcement authority. The proposal aims to balance comprehensive coverage with minimal reporting burden.

[1] At the time of writing this report, the latest available version of the Code of Practice was the third one.

1 Introduction

In this Section, we highlight which parts of the EU AI Act point to reach being considered as a relevant factor when assessing whether a GPAI model presents systemic risks (Section 1.1). Then, we discuss risks created by reach (Section 1.1) and how these risks are reflected in the draft of the GPAI Code of Practice (Section 1.3). This is to show that taking reach into account when assessing whether or not a GPAI model presents systemic risk appears in line with the legal text, and to argue why doing so may be desirable.

1.1 Legal Context

The EU AI Act includes several aspects that relate to the idea of considering risk as a relevant factor for systemic risks:

1. **Recital 110** states that "**Systemic risks** should be understood to **increase with** model capabilities and **model reach**".
2. **Recital 111** states that "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**".
3. **Annex XIII** states that "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; (g) the number of registered end-users**".
4. **Article 3 (65)** states that "'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"

This suggests that the AI Act foresees the possibility of taking reach into account when assessing whether a GPAI model presents systemic risk. To justify why this is reasonable from a risk perspective we turn to the risks created by reach.

1.2 Risks Created by Reach

In this report, we focus on risks created specifically by reach, i.e. risks that are only present when a specific GPAI model is widely used, and that would be absent if the same model was not widely used. The risks from reach that we are concerned with are reflected by the following threat model: GPAI models (e.g. chat assistants) shape the information environment of their users, influencing what information they are exposed to and how. This can have a gradual effect on individual users, their knowledge and beliefs. When a model is widely used, these individual effects add up to a potential effect on society, i.e. systemic risk.

Under this threat model, **risks from model reach mostly correspond to risks from model biases**, in the sense of model tendencies to selectively present information or frame discussions in ways that can subtly shape user perspectives and societal understanding. These biases become particularly problematic when a model achieves widespread adoption, enabling gradual but potentially significant influence over knowledge construction, public discourse, and cultural narratives. Each individual interaction with a biased model might seem inconsequential, but when multiplied across millions of users, these interactions can incrementally shift collective understanding, reinforce existing prejudices, or introduce systematic distortions in how people perceive complex topics.

Examples include:

1. **Political bias** in a language model could influence public discourse and threaten voter autonomy if the model is widely used. For example, a biased model could tend to argue against AI regulation whenever asked about it, which may sway some users' opinion on this issue.
2. **Sociodemographic bias** (e.g. gender or racial bias) in a language model could substantially reinforce and propagate societal prejudices if the model is widely used. This goes beyond domains that are regulated by other parts of the EU AI Act (e.g. hiring, lending, criminal justice). For example, a biased model may reinforce gender stereotypes when completing creative writing tasks.
3. **Cultural bias** (e.g. towards US cultural norms and practices) in a language model could gradually have a homogenising effect on society and result in lower quality of service for less-represented user groups. This too becomes a relevant concern only if the model is widely used.
4. A language model could facilitate the **spread of misinformation** if it is widely used. For example, models could present users with incorrect or misleading information when asked about climate change, which may lead some users to develop misconceptions on this issue.

Importantly, **models do not have to be highly capable (e.g. at the technological frontier) for risks from reach to materialise**. Any reasonably capable language model, for example, can exhibit biases, and if this model is widely used, these biases create risks of harm, as outlined in the examples above.

Note that **risks from reach and frontier capabilities may have distinct risk profiles**. For frontier capabilities, the implicit concern is that the best GPAI models, at some point, will have (catastrophically) dangerous capabilities. For example, a language model may create or facilitate new cybersecurity or CBRN (Chemical, Biological, Radiological and Nuclear) threats if it is sufficiently capable. There is no consensus on the probability of these risks materialising or their exact severity. By contrast, reach alone does not plausibly cause catastrophes, meaning that risks from reach likely have lower severity. However, GPAI models are already very widely used, meaning that risks from reach are almost certainly materialising right now.

Relatedly, **risks from reach materialise from regular/benign use of GPAI models and systems**, whereas capability risks are more closely tied to (although not entirely contingent on) malicious intent or deliberate misuse. This makes risks from reach more pervasive. It also means that **detecting and mitigating risks from reach requires different methods** than detecting and mitigating capability risks.

Notes on scope:

1. We focus only on risks from a *specific* GPAI model being widely used because the EU AI Act is concerned with regulating specific models. There are risks from AI in general being widely used, like the disruption of labour markets. There are also risks of disparate access to AI tools. These issues are outside the scope of this report.
2. Specific GPAI models may also pose systemic risk when deployed in critical contexts (e.g. infrastructure, healthcare, financial systems). This alternative conception of "reach" focuses on deployment context (i.e. where and how the model is used) rather than user volume, and is also outside the scope of this report.

1.3 Risks from Reach in the Code of Practice

We have established that risks from reach are distinct from risks from frontier capabilities. The EU AI Act could plausibly focus on mitigating one or the other, or both. In Section 1.1, we described the legal context. In Section 3 we will outline a concrete approach for taking into account reach when assessing whether a GPAI model presents systemic risks. The missing question we answer now is: **If we were to create a methodology for taking reach into account, and a GPAI model was classified as a GPAI model with systemic risk in part (though not necessarily only) due to its reach, would the obligations triggered by the classification meaningfully address risks from reach?**. In other words, would the obligations faced by the provider actually lead to mitigation of the risks stemming from the reach of the model?.

To address this question, we look to the Safety and Security Section of the 3rd draft of the GPAI Code of Practice (CoP) (Independent Experts, 2025), which outlines obligations for providers of GPAISR models. Most relevant to Question 3 is Appendix 1.1 of the CoP which proposes a Systemic Risk Taxonomy, including sources of risk, as a framework for structuring provider risk assessments and mitigations.

Overall, **reach does not have a prominent role in the CoP risk taxonomy**. The taxonomy draws on the same legal basis we cited in Section 1.1 and mentions reach as a pathway to "significant impact" in that context. However, the four "selected types of systemic risk" that are singled out by the authors are all risks from frontier capabilities rather than reach, which is intentional (p.57: "The list has been sharpened to focus on risks that are specific to the high-impact capabilities of general-purpose AI models."). Appendix 1.2 lists "other types of risks for potential consideration in the selection of systemic risks", which includes reach-related risks such as "risk to democratic processes" and "risk to non-discrimination". Very explicitly, these are framed as secondary considerations. Similarly, Appendix 1.4.1 defines "model propensities, which encompass inclinations or tendencies of a model to exhibit some behaviours or patterns" as potential sources of systemic risk. These propensities include "discriminatory bias". And Appendix 1.4.3. lists "model affordances" that may create systemic risk, including the “number of model business users and number of end-users, including the number of end-users using an AI system in which the model is integrated” Therefore, the CoP acknowledges that some systemic risks may manifest due to reach, independent of frontier capabilities. However, the CoP very clearly does not make risks from reach a major point of emphasis.

Similarly, the obligations for GPAISR models outlined by the CoP make little specific mention of reach or the risks created by reach. Some obligations, such as red-teaming and model weight protections, are primarily relevant to risks created by frontier capabilities, not those created by reach. However, most obligations are sufficiently general to be applied to risks created by reach just as effectively as to any other risk. For example, providers are tasked with creating “Model Reports” to document risk assessments and mitigations. These reports may plausibly include evaluations for bias alongside evaluations for dangerous capabilities. Therefore, **the commitments for GPAISR models providers outlined in the CoP seem compatible with risks resulting from model reach**.

Beyond the considerations raised in the above discussion, the extent to which reach should be considered as a factor when assessing whether a GPAI model presents systemic risks depends on many additional factors, including:

1. **Intended scope:** A criterion based solely on reach would broaden focus from models with frontier capabilities to widely used models with weaker capabilities. This expansion could create compliance burden for a broader set of providers and potentially dilute the Act’s focus via the GPAISR obligations on what some perceive to be the most critical AI risks (i.e. from frontier capabilities).

2. **Political considerations:** Emphasising risks from reach, particularly around bias and discrimination, could make the regulation more contentious in jurisdictions that view such

concerns as primarily social or political rather than technical safety issues.

3. **Regulatory capacity:** Monitoring compliance across both reach and capability dimensions would require expertise in both types of risk. For instance, the enforcement authority would need the capacity to evaluate both traditional safety and security concerns and broader societal impacts, which could potentially strain limited resources and expertise.

Weighing these considerations against each other is beyond the scope of our work in this report.

2 Metrics

In this Section, we provide the definition of "reach" that we then use throughout the rest of this Chapter (Section 2.1). We identify and describe the pros and cons of different metrics for measuring reach, with a view to using these metrics as a factor for determining whether a GPAI model presents systemic risks (Sections 2.2-2.4). We also specifically argue against the usefulness of public information or surveys for measuring model reach (Sections 2.5-2.6). Measuring the reach of GPAI models is difficult, and for many reach metrics, the cons clearly outweigh the pros. In Section 3, we make a concrete proposal for how GPAI model reach could be taken into account as part of assessing whether a GPAI model presents systemic risks.

2.1 Definition of Reach

There is no universal definition of "reach". In the EU AI Act, the reach of GPAI models is most directly tied to their usage, either by "registered business users" or "end users" (see Section 1). Therefore, **we define reach to be concerned with the number of people who directly interact with a given GPAI model**. The key term in our definition is "direct interaction", which we take to mean that users provide inputs to GPAI models, which models respond to. This contrasts with "indirect exposure" to GPAI outputs, where individuals encounter content or decisions influenced by GPAI models without directly engaging with the models themselves.

For example, direct interaction (in scope) would include[2]:

1. A user entering prompts into a chatbot interface.
2. A developer making API calls to a language model.
3. A business user uploading documents for AI analysis.
4. An individual using an AI-powered image generation tool.

By contrast, indirect exposure (out of scope) would include:

1. Reading an AI-generated article without knowing its origin.
2. Being shown AI-curated content recommendations (e.g. social media feeds).
3. Receiving automated decisions based on AI analysis (e.g. spam filters).
4. Encountering AI-generated images in marketing materials.

We exclude indirect exposure because including it would make the scope of reach effectively infinite - every internet user could be considered "in reach" of major GPAI models and systems - rendering reach meaningless for regulatory purposes.

2.2 Access Methods

The metrics we can use to measure GPAI model reach depend on how the GPAI model is made available to users, i.e. the way in which users can directly interact with the model. We see two

[2] In most cases, this direct interaction with the GPAI model will take place through an AI system, including all potential instances. The reach of the model will be linked with the cumulative reach of all AI systems that integrate and provide access to the GPAI model.

primary ways in which GPAI models can be made available to users by model providers: 1) through **user interfaces**, and 2) through **APIs**. On the one hand, user interfaces are usually part of AI systems, so access to the GPAI model will typically be provided through an AI system. On the other hand, APIs enable direct, programmatic access to the GPAI model, which can then be integrated into various AI systems.

Thus, the same GPAI model, open or closed, may be made available to users in different ways, by the same or different providers. This may be done publicly or privately. In the following, we outline potential reach metrics that can be measured by providers depending on how they make models available to users.

Note on terminology: In this report we are using *model provider* to describe the person/organisation making the model available for users to interact with. The model provider may or may not be the *model developer*, i.e. the person/organisation that developed the model. These are intuitive definitions that we use because we find them clearest in the context of this document. For the EU AI Act definitions, see Article 3 (3-4).

Note on private use: We exclude cases where individuals host models exclusively for their own personal use from our consideration, because we do not expect such use to contribute meaningfully to model-specific user counts in the order of magnitude that is relevant for assigning GPAISR status (likely millions of users – see Section 3). We also consider it impractical to require individuals to report this information, or to approximate private use based on public information (see Section 2.5).

Note on geographic location: We understand that, given the scope of the EU AI Act, we would likely have to restrict any reach measurements to reach *within the EU*. Therefore, we use as one criterion for assessing each reach metric how easily it could be restricted to measure reach within the EU.

Note on types of usage: We assume that we cannot know anything about the content of user interactions with GPAI models. Consequently, we do not focus on specific types of interactions. At the most granular level, all reach metrics we discuss in the following concern only whether there is or is not an GPAI model-user interaction.

2.3 Reach Metrics - User Interface Access

A user interface (UI) enables users to directly interact with GPAI models, through a web browser, desktop or mobile application[3]. Examples of public web interfaces for interacting with GPAI models include https://chatgpt.com, https://labs.perplexity.ai/ and https://huggingface.co/chat/.

User interfaces can be hosted either by model developers themselves (like OpenAI's ChatGPT interface) or by third party providers integrating open or closed models. User interfaces may offer users multiple model options to choose from. They can be public, like ChatGPT and HuggingChat, or private, like in the case of company-specific implementations limited to employee use.

User interface access to GPAI models typically requires user registration. For registered users, reach can be measured easily and directly by the following metric:

[3] These examples would likely qualify as AI systems.

Number of registered users sending more than N requests to a specific model within a specific time T

- \+ Easy to track (by organisations hosting the interface)
- \+ Easy to report (by organisations hosting the interface)
- \+ Likely tracked already (by organisations hosting the interface)
- \+ Relatively easy to restrict to EU users (by organisations hosting the interface)
- \+ Little noise: one user = one person in most cases
- ? Need to decide on threshold N for what constitutes "use" of a model
- ? Need to decide on time T (e.g. monthly)

However, not all user interfaces require registration. For non-registered users, reach would have to be measured using a slightly noisier metric:

Number of unique IP addresses sending more than N requests to a specific model within a specific time T

- \+ Easy to track (by organisations hosting the interface)
- \+ Easy to report (by organisations hosting the interface)
- \+ Likely tracked already to some extent (by organisations hosting the interface)
- \+ Relatively easy to restrict to EU users (by organisations hosting the interface)
- – Slightly more noise: same user may use multiple IPs, VPNs, etc.
- ? Need to decide on threshold N for what constitutes "use" of a model
- ? Need to decide on time T (e.g. monthly)

Both metrics measure reach almost exactly as we define it. That is, how many people directly interact with a given model during a specific time frame.

Overall, **whoever makes GPAI models available via a user interface will have or can easily obtain a good estimate of how many people use which model in what time frame**.

2.4 Reach Metrics - API Access

An API (Application Programming Interface) is a set of protocols and tools that allows software applications to communicate with GPAI models programmatically. Examples of public APIs for interacting with open and/or closed GPAI models include those offered by OpenAI (OpenAI, 2025), Amazon Bedrock (AWS, 2025), Hyperbolic (Hyperbolic, 2025) and Replicate (Replicate, 2025).

Public APIs are either offered by model developers directly (like the OpenAI API) or by third party providers who offer access to open models (e.g. Replicate offering access to Llama models from Meta) or closed models (like Amazon Bedrock offering access to Claude from Anthropic). APIs can also be private, like in the case of companies making models available to their employees. Making calls to an API almost always requires an API key, which is associated with a user account.

Measuring reach in the sense that we define it (i.e. how many people use a specific model, see Section 2.1) is more complicated when GPAI models are offered through APIs. We believe the following metric to be the most plausible option:

Number of users making more than N API requests to a specific model within a specific time T

- + Easy to track (by organisations hosting the API)
- + Easy to report (by organisations hosting the API)
- + Likely tracked already (by organisations hosting the API)
- + Relatively easy to restrict to EU users (by organisations hosting the API)
- **?** Need to decide on threshold N for what constitutes "use" of a model
- **?** Need to decide on time T (e.g. monthly)

This metric is preferable to tracking the number of API keys because API keys are not associated with specific models, and single users may generate many keys. This metric is also preferable to tracking the number of API calls for a specific model because single users may make many calls.

However, while individual users (e.g. researchers) may interact directly with models via an API, APIs are often used to power user interfaces, which may then in turn be used by many users. This means that one "model user" (as identified by the API provider) may correspond to many more actual model users. Therefore, **counting only how many users make calls to a specific model via an API would likely underestimate the actual reach of the model via the API, to a degree that is difficult to quantify, but may at times be substantial**.

In our concrete proposal for a reach trigger for GPAISR status (Section 3), we address this limitation by combining API and user interface metrics.

2.5 Measuring Reach using Public Information

All reach metrics we discussed above are based on usage information that is available only to model providers, and very difficult, if not impossible, to reconstruct by third parties. This may make it tempting to look towards public information on GPAI models as proxy measures for actual usage / model reach. However, **we believe all plausible reach metrics based on public information to be inadequate because they are too disconnected from model usage**.

For example, it would be relatively easy for anyone to track the following:

1. Hugging Face model downloads or likes.
2. GitHub stars or forks on the model repository.
3. Social media mentions of the model's name.
4. Google search volume for terms related to the model.
5. Academic paper citations referencing the model.
6. Developer forum discussions/threads about the model.
7. News article coverage frequency and sentiment.

These metrics may measure model popularity in one sense or the other, but **they cannot measure with any reasonable accuracy how many people use a specific model within a specific time frame**. To assign GPAISR status based on reach, we need to know when specific models surpass a usage threshold, and when they fall below this threshold again. We need to measure usage restricted to the EU. None of the public proxy metrics listed above allow for this, and we strongly believe that this excludes these metrics from consideration, before even mentioning other, more metric-specific downsides, such as their lack of applicability to closed models (e.g. in the case of Hugging Face or GitHub metrics), or how easily they could be gamed (e.g. in the case of forum activity or social media mentions).

2.6 Measuring Reach using Surveys

Another potential approach for measuring GPAI model reach is the use of surveys. In principle, the enforcement authority could run regular representative surveys across EU member states, asking people about which GPAI models they use. The main benefit of this would be that it cuts out the need for providers to cooperate and report usage metrics. However, **we believe survey-based reach measurement to be inadequate** for two main reasons:

1. The Chapter V of the EU AI Act targets GPAI models (e.g. gpt-4o-2024-08-06) rather than systems (e.g. ChatGPT), and regular users are often not aware which model, let alone which model version, they are interacting with. A survey may enable reasonable estimates of EU user numbers for systems or model families, but this is not useful information for deciding on GPAISR at the model level. Model providers, by contrast, (can) know exactly which model they are deploying as part of any interface or API that they make accessible to users.
2. Even if survey participants were able to identify which exact models they used, and if we assumed relatively little noise in this self-reporting, the enforcement authority would still need to gather this information very regularly and from a large and representative sample of people, creating substantial financial and time costs. This is because GPAISR status based on reach would require regular (e.g. monthly) re-assessment, as models become more (or less) widely used. By contrast, model providers can be assumed to mostly track model usage already, and once reporting is set up, there is little marginal cost on either the EU or the provider side in regular reporting.

Overall, we therefore believe that **accurate measurement of model reach requires reach metrics to be reported by model providers**. This is the foundational assumption for our concrete proposal in the next Section. We do not address in this report the question whether or not such reporting is within the scope of the AI Act.

3 Concrete Proposal for the Design of a Reach Criterion

Our goals in designing a criterion based on reach that could be used to inform, amongst other criteria, whether a GPAI model should be a GPAISR model are to:

1. **Maximise coverage across different GPAI models**, so as not to disproportionately create obligations for some model types over others. Specifically, we want to:
   (a) Capture the usage of both open and closed models.
   (b) Capture the usage of both publicly and privately hosted models.
2. **Minimise reporting cost and complexity**, so as not to overburden reporting and monitoring stakeholders.

We propose that the reporting burden falls only on model providers, i.e. the person/organisation making the model available for users to interact with. As mentioned above, we do not address the question of whether or not such reporting is in scope of the AI Act.

For each GPAI model that they make accessible, providers would report the number of users interacting with the model through an API and/or user interface, each month. These metrics are likely already tracked by most providers.

Only providers serving many users would need to make these reports. This overall reporting threshold could then be calibrated.

Then, the enforcement authority could aggregate the reported metrics across providers per model/month and compare the total user count to a set (possibly calibrated) threshold for assessing the extent to which the reach of the model can be viewed to contribute to its systemic risks.

3.1 Overall Reporting Thresholds for Model Providers

A model provider is a person/organisation making a GPAI model available for users to interact with, either through an API or a user interface (UI). Providers may offer access to open or closed models, either publicly (e.g. on the web) or privately (e.g. on company intranet). Providers may be model developers (e.g. OpenAI) or provide access to third-party models (e.g. Replicate). Providers may be very large (e.g. OpenAI, Amazon) or small (e.g. a start-up), and serve a small or large number of users.

For the purposes of using reach as a factor when assessing the systemic risks presented by a GPAI model, we want to measure model usage as accurately as possible. However, requiring every provider, no matter how small they are or how few users they serve, to report user counts to the enforcement authority would be unreasonable (separately to the question of whether or not this is in scope of the AI Act). Therefore, we need to set an **overall reporting threshold** below which providers would not be required to report user counts.

We believe that the most sensible metric for setting an overall reporting threshold is:

Number of users making more than N requests (via API or interface) to any GPAI model within a specific time T (e.g. monthly)

+ Easy to track (by provider)
+ Easy to report (by provider)
+ Likely tracked already (by provider)
+ Relatively easy to restrict to EU users (by provider)

Compared to the reach metrics in Section 2, this metric does not differentiate between users of different GPAI models, and it combines usage via API and user interface, to simplify reporting / tracking.

This metric is preferable to metrics based on provider size (e.g. number of employees or revenue) because even small providers may serve many users.

The threshold should be set so that most providers know their reporting status without needing to do much analysis. It should be obvious to providers with few users that they are not in scope, and vice versa.

There should likely also be some overhang to reporting requirements, i.e., once considered in-scope, providers would have to report reach metrics until they can show that they no longer meet the overall reporting threshold.

3.2 Reach Metrics to be Reported

We assume that providers know which of the models they make accessible are GPAI. For each GPAI model, providers would have to report the following:

— If the model is made available through one or multiple **user interfaces**:

1. For registered users: Number of registered users sending more than N requests to the model within a specific time T.
2. For non-registered users: Number of unique IP addresses sending more than N requests to the model within a specific time T.

— If the model is made available through one or multiple **APIs**:

3. Number of users making more than N API requests to the model within a specific time T (assuming that API usage requires registration).

The **rationale** is the following: As outlined in Section 2, reach metrics for user interfaces are easily recorded and allow us to track reach as we define it (i.e. direct user interactions). Reach metrics for models made accessible via API, on the other hand, likely underestimate user counts because APIs may power user interfaces with many more users. We address this shortcoming by combining user interface metrics with API metrics. If a single API user corresponds to many more users of a user interface, then the provider of that user interface (who may or may not be the API provider) would have to report user counts for this interface.

There are plausible scenarios that lead to double counting of users:

— The same users may use the same model from multiple providers.

— The same user may use the same model from the same provider through both their user interface and their API (without the provider correcting for this).

— API "users" that power user interfaces would be counted once as an API user on top of the user count of the user interface.

However, the likely overall impact of this on user counts seems negligible.

For more efficient reporting, there should likely be **additional reporting thresholds for web interfaces and APIs**, so that providers do not need to record user counts for web interfaces and APIs that are used by very few users.

The **reporting format** could be a standardised CSV file, where each line corresponds to a GPAI model, and columns correspond to metrics 1-3 above. Providers would have to share such a CSV file with the enforcement authority every month, for example through a web portal. An API could also be established to facilitate easier reporting.

The reported metrics could then be aggregated across providers on a per-model, per-month basis, and the total user count could then be used to assess the extent to which the reach of the model can be viewed to contribute to its systemic risks.

3.3 Calibrating the Reach Criterion

There are several parameters in our proposed reach criterion which can be used for calibration. Two thresholds are particularly impactful:

1. The **reporting threshold** based on how many users interact with the GPAI models made accessible by a single provider. This impacts how many providers would need to report reach metrics in the first place.
2. The **"trigger" threshold** based on how many users across all providers interact with a given GPAI model. This impacts when the reach of the model can be considered as a relevant factor contributing to its systemic risks, and therefore, how many additional GPAI models might be subject to GPAISR obligations.

Deciding on these two thresholds is difficult because, until reporting requirements are implemented, **it is unclear how many providers there are, how many users they each serve, and how many users overall use a given model**. There is very little public information on this, and even less that is restricted to EU users.

For providers, our expectation is a power-law distribution, where the top providers serve most users, and a long tail of providers serves few users each. The reporting threshold could be adjusted over time to capture a certain proportion of overall usage (e.g. 90% of users) or a certain number of providers (e.g. top 100 providers). We expect a plausible reporting threshold to be in the thousands of users, resulting in hundreds of providers that are required to report metrics to the enforcement authority.

For models, we also expect a power-law distribution, where the top models have many millions of users, and a long tail of models has few users each. The "trigger" threshold should likely be adjusted to capture only the top models with the most reach. For example, we could consider reach to be a relevant factor for systemic risks only when it affects millions of users.

To solve the cold start problem, reporting requirements could be activated several months before exact thresholds are decided and obligations are triggered. Then, the reporting threshold could be calibrated to balance coverage and reporting burden.

At a more granular level, another parameter that will require calibration is **the number of requests N that constitutes model usage (via API or interface) in a given time**. We believe this should be higher than 1 to avoid counting one-off interactions, but not very high, so that casual use is still captured.

Finally, the **reporting frequency** could also be adjusted. We believe monthly reporting to be sufficiently granular, and a natural fit with how providers track usage.

4 Conclusions

In this report, we examined the feasibility and desirability of using reach as a factor when assessing whether a GPAI model presents systemic risks under the EU AI Act. Our analysis suggests that taking reach into account may have a legal basis within the Act's provisions. We also argued that reach creates distinct risks that are largely independent of frontier capabilities and require different mitigation approaches. In particular, we argued that risks from reach primarily stem from model biases that can gradually influence societal understanding and discourse when models achieve widespread adoption. These biases become particularly problematic at scale, where individual interactions with biased models may seem inconsequential but can collectively shift public opinion, reinforce existing prejudices, or introduce systematic distortions in how people perceive complex topics.

From a technical implementation perspective, we showed that measuring model reach through provider-reported usage metrics is both feasible and practical. Our proposed framework, which combines user interface and API metrics while establishing appropriate reporting thresholds, would enable sufficiently comprehensive coverage of model usage while minimising administrative burden. The aggregation of these metrics by the enforcement authority would provide a reliable basis for assessing the reach of a GPAI model, and determining the extent to which this reach may be contributing to the systemic risks of the model.

The desirability of considering reach as a factor for determining whether a GPAI model should be classified as a GPAISR model, however, depends on broader policy considerations regarding the intended scope of the EU AI Act, regulatory capacity, and political priorities. A reach trigger could broaden the focus beyond frontier models to include widely used models with weaker capabilities, potentially creating compliance burdens for a broader set of providers while requiring regulators to develop expertise in both narrow frontier safety concerns and broader societal impacts. These trade-offs merit careful consideration in determining whether and how reach should be taken into account in assessing whether a GPAI model presents systemic risks.

References

AWS, 'Amazon Bedrock API Reference'.
https://docs.aws.amazon.com/bedrock/latest/APIReference/welcome.html, 2025.

European Commission, 'Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689 (AI Act)'.
https://ec.europa.eu/newsroom/dae/redirection/document/118340, 2025.

European Union, 'Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence'.
https://eur-lex.europa.eu/eli/reg/2024/1689/oj, 2024.

Hernández-Orallo, J., Sevilla, J., Gómez, E. and Fernández-Llorca, D., 'General-Purpose AI Models in the AI Act: Capabilities, Generality, Systemic Risks and Compute', European Commission, Joint Research Centre, JRC139341, 2024.

Hyperbolic, 'Explore Models'. https://app.hyperbolic.xyz/models, 2025.

Independent Experts, 'Third Draft of the General-Purpose AI Code of Practice'.
https://digital-strategy.ec.europa.eu/en/library/third-draft-general-purpose-ai-code-practice-published-written-independent-experts, 2025.

OpenAI, 'OpenAI developer platform'. https://platform.openai.com/, 2025.

Replicate, 'Run AI with an API'. https://replicate.com/, 2025.

List of abbreviations and definitions

AI Artificial Intelligence

API Application Programming Interface

CBRN Chemical, Biological, Radiological and Nuclear

CSV Comma Separated Values

GPAI General-Purpose Artificial Intelligence

GPAI models General-Purpose Artificial Intelligence models

GPAISRs/GPAISR General-Purpose Artificial Intelligence models with Systemic Risk

IP Internet Protocol (address)

UI User Interface

VPN Virtual Private Network

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

Scan the QR code to visit:

The Joint Research Centre: EU Science Hub
https://joint-research-centre.ec.europa.eu

Publications Office of the European Union