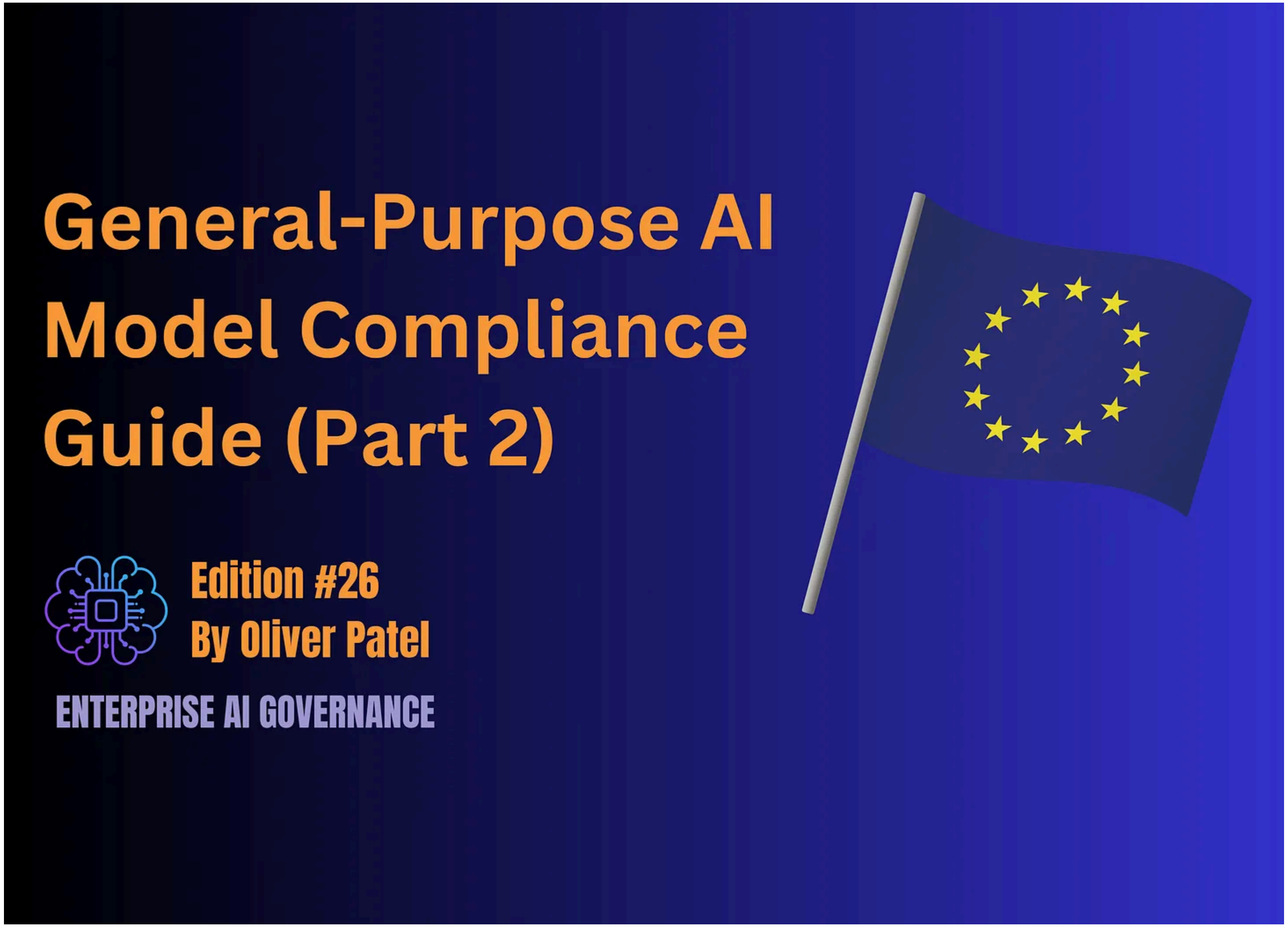


Hey 👋

I'm Oliver Patel, author and creator of ***Enterprise AI Governance***.

This free newsletter delivers practical, actionable, and timely insights for AI governance professionals.

My goal is simple: to empower you to understand, implement, and master AI governance.

If you haven't already, sign up below and share it with your colleagues. Thank you!

Follow me on LinkedIn for more frequent updates.

Welcome to Part 2 of the General-Purpose AI (GPAI) Model Compliance Guide. This 3-part series is posted exclusively on Enterprise AI Governance. To all subscribers and new readers, thanks for supporting the newsletter!

This week's edition provides a comprehensive yet accessible overview of the EU AI Act's provisions for **GPAI Models with Systemic Risk**.

You will learn:

✅ What is a GPAI model with systemic risk?
✅ What are the notification and exception procedures for providers of GPAI models

with systemic risk?
✅ What are the compliance obligations for providers of GPAI models with systemic risk?
✅ What exactly is a "systemic risk"?
✅ Deep dive on the *GPAI Code of Practice: Safety and Security Chapter*

If you haven't read Part 1, you should check it out here and read it first. It provides a detailed breakdown of the **Obligations for Providers of GPAI Models**, including what the core obligations for all GPAI models are, how and when these obligations will be enforced by the AI Office, as well as specific considerations for open-source models and legacy GPAI models (released before 2 August 2025). All this is essential background information that is necessary to fully understand the compliance implications for GPAI models with systemic risk.

Part 3, coming next week, will cover the knotty issue of **'Downstream Actors': Modification, Deployment, and Use of GPAI Models**. This will address the important question of who exactly is a provider of a GPAI model—it could be you!

Note: *this series assumes familiarity with the EU AI Act, its core concepts, and the topic of general-purpose AI and foundation models more broadly. Also, I was not involved in the multi-stakeholder process of drafting and developing the EU's GPAI Code of Practice. Finally, none of this should be taken as legal advice. Always consult a legal professional.*

The following official sources have been used to create this guide:

- EU AI Act full text
- European Commission Guidelines for Providers of GPAI Models
- GPAI Code of Practice
  - Transparency Chapter
    - Model Documentation Form
  - Copyright Chapter
  - Safety and Security Chapter
- Template for publishing GPAI Model Training Data Summary

What is a GPAI model with systemic risk?

There are two broad categories of GPAI models that the AI Act regulates. These are:

1. GPAI models
2. GPAI models with systemic risk

As explained in Part 1 of this series, GPAI models *"are trained with a large amount of data using self-supervision at scale [...] display significant generality and are capable of performing a wide range of distinct tasks"*. The European Commission's recent guidelines also clarify that if an AI model's training compute exceeds 10^23 floating-point operations, and it can generate language (either text or audio), or generate image or video based on text inputs, then it should be considered a GPAI model.

An AI model is classified as a GPAI model with systemic risk if it meets the above definition and criteria and also has "**high impact capabilities**". The AI Act defines this as capabilities that "*match or exceed the capabilities recorded in the most advanced GPAI models*".

If the cumulative amount of computation used to train a GPAI model exceeds 10^25 floating-point operations, then it is, by default, presumed to have "high impact capabilities" and thus classified as a GPAI model with systemic risk.

By relying on this compute threshold, the EU's position is that there is a direct correlation between how much computational resource is used to train an AI model and both the general-purpose capabilities of the model and the level of risk that it poses.

This means that the larger an AI model is (e.g., in terms of number of parameters) and the more data that is used to train it (e.g., in terms of different examples), the more likely it is to be classified as a GPAI model with systemic risk.

Whilst the European Commission acknowledges that "*training compute is an imperfect proxy for generality and capabilities*", it argues that it is "*the most suitable approach at present*".

However, the European Commission is empowered to amend this compute threshold, or introduce an entirely new indicator, via a delegated act. This means it can do so independently, without reopening and amending the AI Act itself. However, the European Parliament and the Council (i.e., the member states) have a right to object to any such changes.

Interestingly, it is possible for a GPAI model to be classified as a GPAI model with systemic risk even if it does not meet the 10^25 floating-point operations compute threshold. This would require the European Commission to determine that it nonetheless has high impact capabilities, despite the lower cumulative amount of compute used to train it.

In making such a decision—that would prove controversial due to the impact on the impacted provider—the European Commission would consider factors like the size of the model, input and output modalities, benchmark and evaluation results, model autonomy level, and the number of end-users. Ultimately, it would have to prove that its capabilities match or exceed those of the most advanced GPAI models, despite the fact that less compute was used to train it.

What are the notification and exception procedures for providers of GPAI models with systemic risk?

The key point for enterprises is that they must carefully forecast, measure, track, and record their estimates of the amount of computational resource used to develop, train, modify, and fine-tune GPAI models, in order to determine what compliance obligations they may have to adhere to.

When estimating and measuring compute levels, the European Commission's guidance is that providers should "*as a general rule, account for all compute that contributed or will contribute to the model's capabilities*". This even includes the compute expended to generate synthetic data for training, even if not all the synthetic data was eventually used to train the GPAI model.

Once an organisation knows that a GPAI model it has developed (or is in the process of developing) meets the threshold for training compute (which means it is classified as a GPAI model with systemic risk), it must notify the European Commission of this as soon as possible, and within two weeks at the latest. In some cases, this notification will be required before the overall training process is completed (e.g., if the threshold is exceeded mid-training run). This notification should include both the precise computation amount as well as a detailed explanation of how this has been estimated.

In its guidelines, the European Commission recommends that "*providers should estimate the cumulative amount of training compute that they will use*" before the training process begins. If their pre-training estimate surpasses the systemic risk threshold, they should inform the Commission of this.

Zooming out, this notification procedure enables the European Commission's AI Office to fulfill its role as the regulator overseeing and enforcing the AI Act's provisions on GPAI models. It will also promote transparency, as the European Commission will publish a list of all GPAI models with systemic risk that are in scope of the AI Act.

Finally, it is possible for a provider of a GPAI model that is by default classified as a GPAI model with systemic risk to secure an exception. To do this, the provider must demonstrate that its GPAI model does not have "high impact capabilities" and therefore does not pose systemic risks and should not be classified as such, despite surpassing the cumulative compute for training threshold of 10^25 floating-point operations.

Providers can do this by pointing to evidence like benchmark and evaluation results, especially if these demonstrate a capability gap between their model and the most advanced AI models. It is important to note that providers cannot get out of the GPAI model with systemic risk classification merely by implementing robust controls and safeguards which mitigate the systemic risk. To secure an exception, they must convince the European Commission, with cold, hard evidence, that the model genuinely does not have high impact capabilities.

The AI Act describes this as an "exceptional" scenario, requiring European Commission approval. In such instances, the burden of proof will be on the provider.

Given the extensive additional compliance obligations for providers of GPAI models with systemic risk (as compared to GPAI models), the question of which GPAI models are and are not classified as posing systemic risk is significant.

Precisely what these additional compliance obligations are is explained below.

What are the compliance obligations for providers of GPAI models with systemic risk?

Part 1 of this series provides a detailed breakdown of the compliance obligations for providers of GPAI models. In summary, the four core obligations for GPAI models are:

1. Develop, maintain, and keep up-to-date comprehensive technical documentation.
2. Produce and make publicly available a detailed summary of the content and data used to train the GPAI model.
3. Implement a policy to comply with EU copyright and intellectual property law.
4. Cooperate with the European Commission and regulatory authorities and appoint an EU-based authorised representative (if based outside of the EU).

Providers of open-source GPAI models are exempt from obligations 1 and 4. This means they still need to publish a training data summary and implement a copyright compliance policy.

Providers of GPAI models with systemic risk must comply with all of the above obligations. Also, providers of open-source GPAI models with systemic risk are **not exempt** from any of the above obligations. This means that sufficiently advanced and capable open-source AI models are treated the same, from an AI Act compliance perspective, as proprietary models.

In other words, providers of any GPAI model with systemic risk, that is placed on the market or made available in the EU, irrespective of whether it is open-source, must comply with all the above obligations (for providers of GPAI models), as well as the additional obligations for providers of GPAI models with systemic risk.

There are **four core additional obligations** that only apply to providers of GPAI models with systemic risk. These are:

1. Perform model evaluation using state of the art tools and protocols. This includes conducting adversarial testing to enable the identification and mitigation of "**systemic risks**".
2. Assess and mitigate potential systemic risks that may stem from the development, deployment, or use of the GPAI model with systemic risk.
3. Track, document, and report information about serious incidents and any corrective measures to address them.
4. Ensure an "adequate level of cybersecurity protection" for both the GPAI model with systemic risk and the physical infrastructure of the model.

These obligations reflect the fact that EU lawmakers deem it both appropriate and necessary for the most advanced and capable foundation models to be subject to rigorous governance—including stringent safety and security testing and evaluation procedures, the implementation of technical guardrails and safeguards to mitigate risk, continuous monitoring and oversight, and documented accountability and risk ownership—due to the widespread use of these models and the distinct possibility that this use could lead to significant negative impact.

However, the text of the AI Act itself does not provide much detail about how to approach and implement the above four obligations. That is why there is a GPAI Code of Practice. The Code provides a detailed, standardised, and step-by-step compliance framework for GPAI model providers.

The *GPAI Code of Practice: Safety and Security Chapter*, which is most relevant for these obligations, is analysed below. But first, we explore the definition of "systemic risk", which is at the heart of these obligations.

What exactly is a "systemic risk"?

The overarching purpose of the above obligations is for providers to uncover and mitigate the systemic risks which their most capable and advanced GPAI models pose. This includes reducing the likelihood of these risks materialising and reducing their impact if they do materialise.

This raises an important question for GPAI model providers: what exactly is a "systemic risk"?

The GPAI Code of Practi*ce* builds on the AI Act by providing additional detail regarding precisely how providers should define, identify, and evaluate the systemic risks that their GPAI models could pose.

It classifies the following four risks as systemic risks. This means that if any providers that are also signatories identify any of these risks, they must be classified and treated as systemic risks:

- **Chemical, biological, radiological, and nuclear (CBRN)**, e.g., a GPAI model that makes it easier or otherwise enables the design, development, and use of CBRN-related weapons or materials.
- **Loss of control**, e.g., a GPAI model that autonomously self-replicates and creates new, more advanced AI models, without human awareness or control.
- **Cyber offence**, e.g., a GPAI model that can be used to significantly lower barriers to entry for scaling cyber attacks.
- **Harmful manipulation**, e.g., a GPAI model that targets large populations of people and uses deceptive techniques to promote harmful or destructive behaviour.

Recital 110 of the AI Act complements this, by providing an illustrative list of examples of systemic risks:

- Major accidents.
- Disruptions of critical sectors.
- Serious consequences to public health and safety.
- Negative effects of democratic processes.
- Negative effects on public or economic security.
- The dissemination of illegal, false, or discriminatory content.

More broadly the GPAI Code of Practice clarifies the essential characteristics of a systemic risk, to further enable their identification. This clarification is based on the formal definition of systemic risk provided in the AI Act. The three essential characteristics of a systemic risk are:

1. The risk is directly related to the GPAI model's high-impact capabilities.
2. The risk has a significant impact on the EU due to its reach or due to the actual or potential negative impact on public health, safety, public security, fundamental rights, or society as a whole.
3. The impact can spread widely, at scale, through connected AI systems and the AI and industry ecosystem more broadly.

The EU's view is that as model capabilities and model reach increase, so do the potential systemic risks. Recital 110 of the AI Act also highlights that such systemic risks can arise due to various factors and causes, including (but not limited to):

- Model misuse
- Model reliability
- Model fairness
- Model security
- Model autonomy level
- Tool access
- Model modalities
- Release and distribution mechanisms
- Potential to remove model guardrails

The detailed information provided in the AI Act and the Code of Practice is sufficient to enable providers of GPAI models with systemic risk to fulfil their obligation of identifying, evaluating, and mitigating the specific systemic risks their GPAI models may pose.

Deep dive on the *GPAI Code of Practice: Safety and Security Chapter*

The final section of this article summarises and analyses the key elements of the *GPAI Code of Practice: Safety and Security Chapter.*

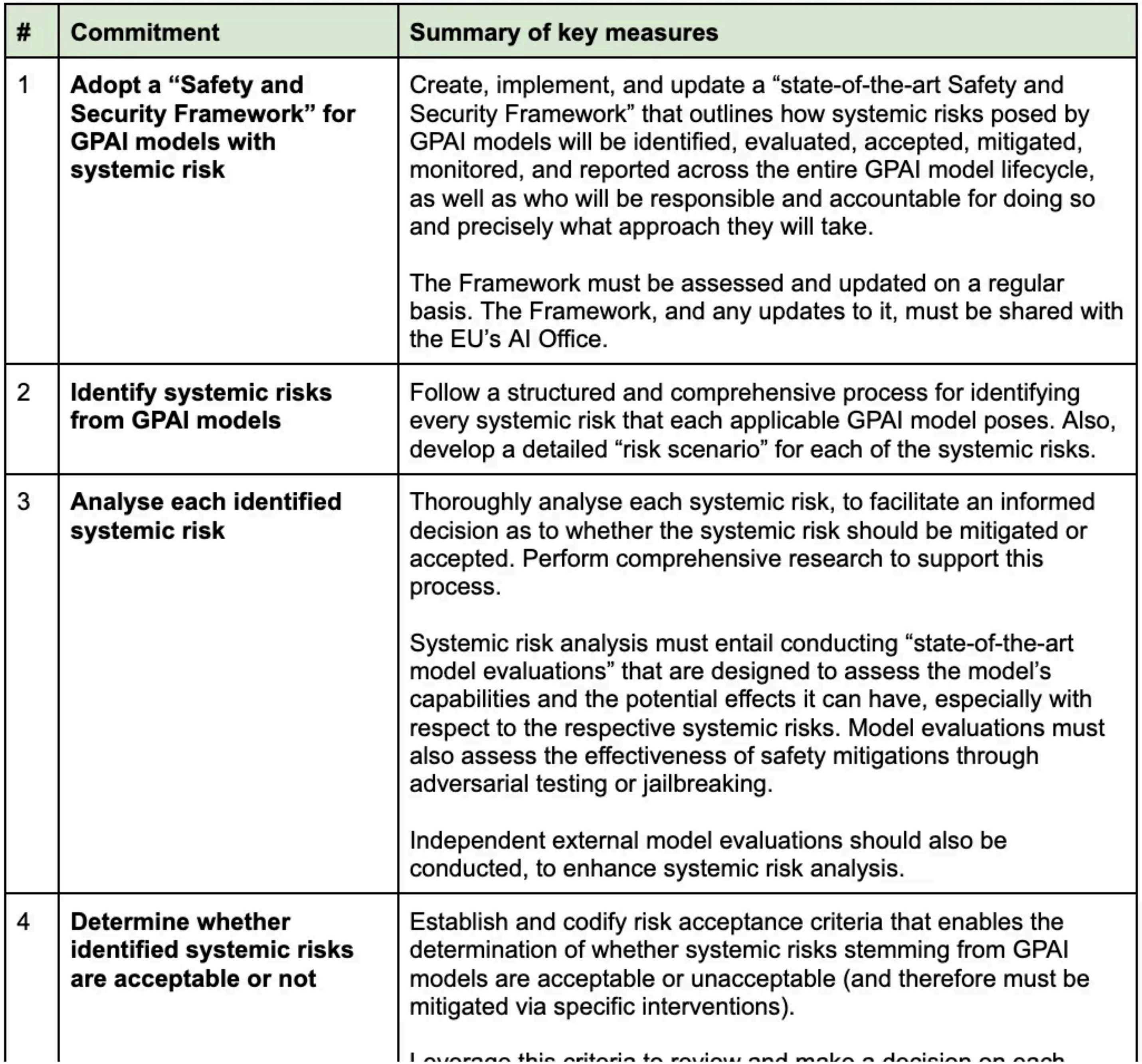
This Chapter focuses exclusively on the specific obligations for providers of GPAI models with systemic risk. It provides a comprehensive and standardised set of commitments and measures that signatory organisations will implement, in order to adhere to the four compliance obligations for GPAI models with systemic risk (detailed above).

For a dedicated explainer on the GPAI Code of Practice, check out this previous post on Enterprise AI Governance. The most important things to know about the GPAI Code of Practice are that i) it was approved by the EU on 1 August 2025, ii) it is a voluntary resource which helps providers comply with the full suite of obligations for GPAI models, and iii) it consists of three chapters: 1) Transparency, 2) Copyright, and 3) Safety and Security.

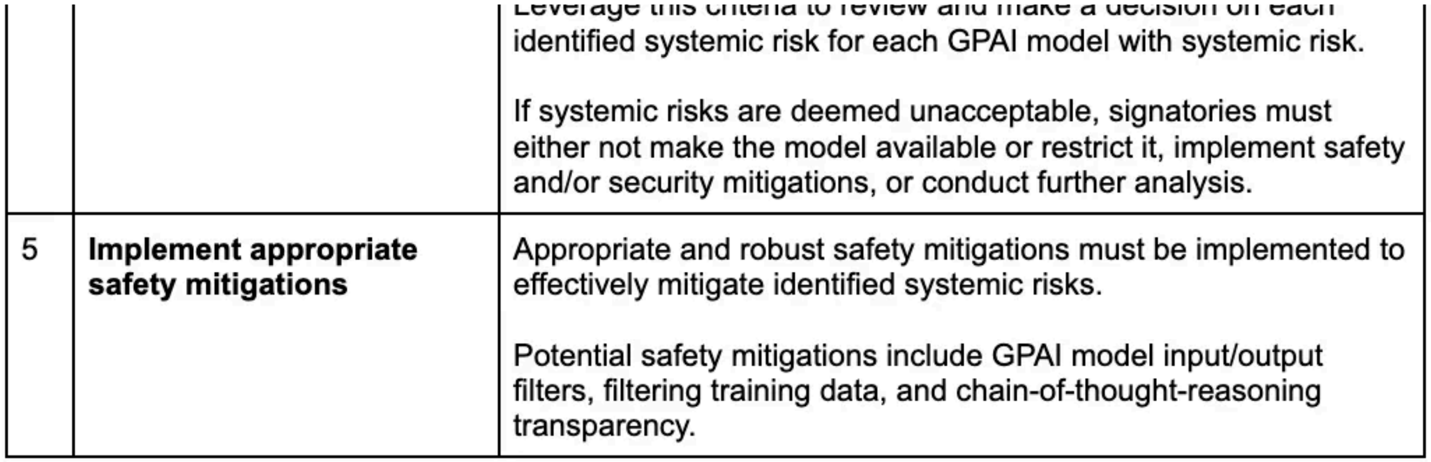
The European Commission strongly encourages providers to sign the Code of Practice and even indicated that it will be more trusting of signatory organisations. However, the Code itself is not law.

The Safety and Security Chapter, which is by far the most detailed of the three chapters, consists of ten commitments. All commitments directly relate to GPAI model with systemic risk risk identification, management, mitigation, treatment, monitoring, ownership, and accountability.

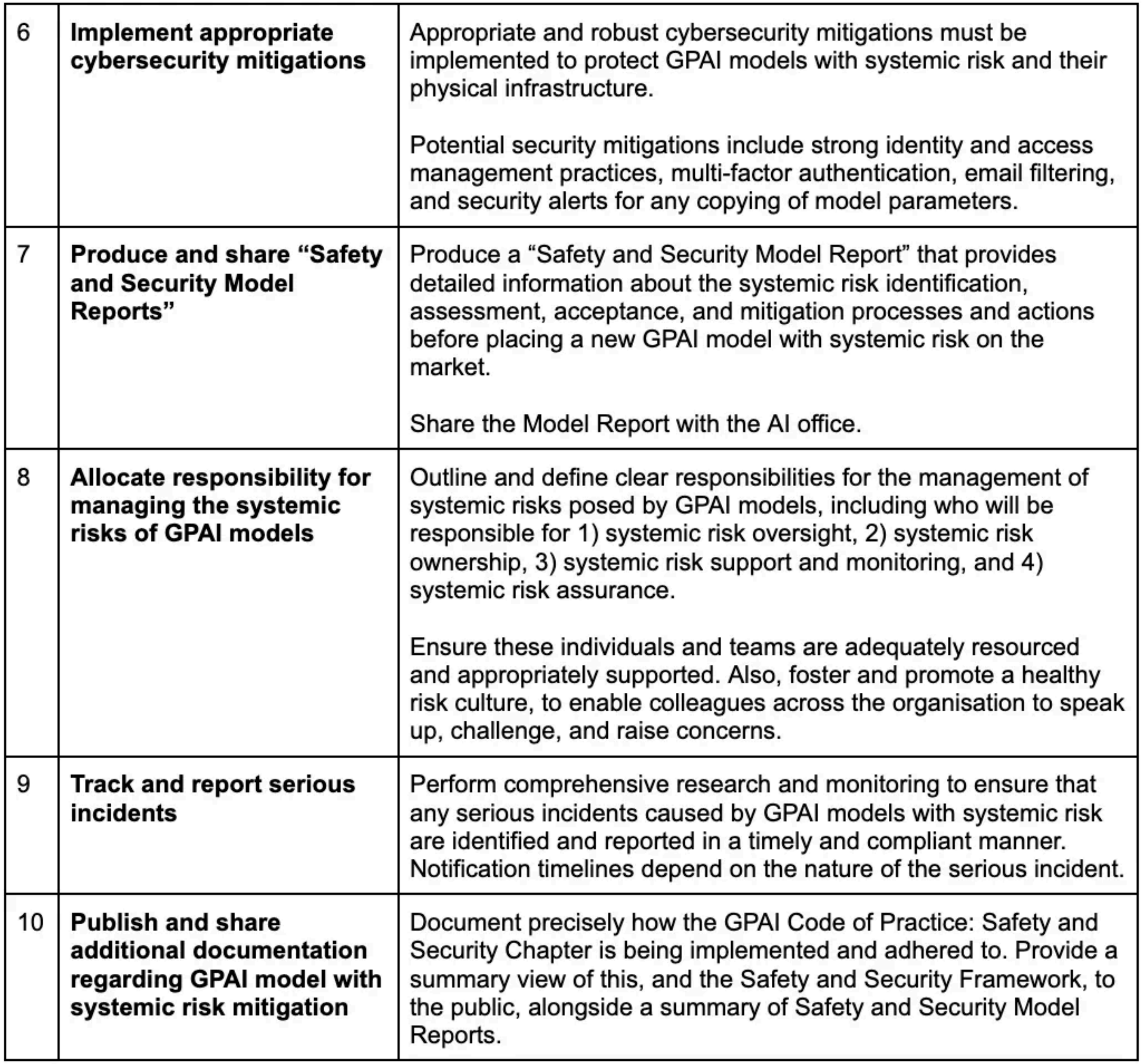
Below is a detailed breakdown of each of the ten commitments and a summary of the most important measures that signatory organisations have committed to implementing.



#	Commitment	Summary of key measures
1	**Adopt a “Safety and Security Framework” for GPAI models with systemic risk**	Create, implement, and update a “state-of-the-art Safety and Security Framework” that outlines how systemic risks posed by GPAI models will be identified, evaluated, accepted, mitigated, monitored, and reported across the entire GPAI model lifecycle, as well as who will be responsible and accountable for doing so and precisely what approach they will take. The Framework must be assessed and updated on a regular basis. The Framework, and any updates to it, must be shared with the EU’s AI Office.
2	**Identify systemic risks from GPAI models**	Follow a structured and comprehensive process for identifying every systemic risk that each applicable GPAI model poses. Also, develop a detailed “risk scenario” for each of the systemic risks.
3	**Analyse each identified systemic risk**	Thoroughly analyse each systemic risk, to facilitate an informed decision as to whether the systemic risk should be mitigated or accepted. Perform comprehensive research to support this process. Systemic risk analysis must entail conducting “state-of-the-art model evaluations” that are designed to assess the model’s capabilities and the potential effects it can have, especially with respect to the respective systemic risks. Model evaluations must also assess the effectiveness of safety mitigations through adversarial testing or jailbreaking. Independent external model evaluations should also be conducted, to enhance systemic risk analysis.
4	**Determine whether identified systemic risks are acceptable or not**	Establish and codify risk acceptance criteria that enables the determination of whether systemic risks stemming from GPAI models are acceptable or unacceptable (and therefore must be mitigated via specific interventions). Leverage this criteria to review and make a decision on each



		Leverage this criteria to review and make a decision on each identified systemic risk for each GPAI model with systemic risk. If systemic risks are deemed unacceptable, signatories must either not make the model available or restrict it, implement safety and/or security mitigations, or conduct further analysis.
5	**Implement appropriate safety mitigations**	Appropriate and robust safety mitigations must be implemented to effectively mitigate identified systemic risks. Potential safety mitigations include GPAI model input/output filters, filtering training data, and chain-of-thought-reasoning transparency.



6	**Implement appropriate cybersecurity mitigations**	Appropriate and robust cybersecurity mitigations must be implemented to protect GPAI models with systemic risk and their physical infrastructure. Potential security mitigations include strong identity and access management practices, multi-factor authentication, email filtering, and security alerts for any copying of model parameters.
7	**Produce and share “Safety and Security Model Reports”**	Produce a “Safety and Security Model Report” that provides detailed information about the systemic risk identification, assessment, acceptance, and mitigation processes and actions before placing a new GPAI model with systemic risk on the market. Share the Model Report with the AI office.
8	**Allocate responsibility for managing the systemic risks of GPAI models**	Outline and define clear responsibilities for the management of systemic risks posed by GPAI models, including who will be responsible for 1) systemic risk oversight, 2) systemic risk ownership, 3) systemic risk support and monitoring, and 4) systemic risk assurance. Ensure these individuals and teams are adequately resourced and appropriately supported. Also, foster and promote a healthy risk culture, to enable colleagues across the organisation to speak up, challenge, and raise concerns.
9	**Track and report serious incidents**	Perform comprehensive research and monitoring to ensure that any serious incidents caused by GPAI models with systemic risk are identified and reported in a timely and compliant manner. Notification timelines depend on the nature of the serious incident.
10	**Publish and share additional documentation regarding GPAI model with systemic risk mitigation**	Document precisely how the GPAI Code of Practice: Safety and Security Chapter is being implemented and adhered to. Provide a summary view of this, and the Safety and Security Framework, to the public, alongside a summary of Safety and Security Model Reports.