THE FUTURE SOCIETY

December 2023

EU AI Act Compliance Analysis

General-Purpose AI Models in Focus

Contact: info@thefuturesociety.org

Cite as: “EU AI Act Compliance Analysis: General-Purpose AI Models in Focus” (The Future Society, December 2023). https://thefuturesociety.org/EU-AI-Act-Compliance-Analysis

Table of Contents

THE
FUTURE
SOCIETY

EU AI Act Compliance Cost Analysis

General-Purpose AI Models in Focus

December 2023

Key takeaways:

- Compliance costs for GPAI models even for the most heavy-handed set of requirements & obligation (EP's Article 28b) are negligible, despite very conservative assumptions.
- Specifically, **compliance costs represent 0.07% to 1.34% of Total Investment** per model, for building models between 10^24 and 10^26 FLOP.
- We **confidently recommend more stringent binding requirements than what is currently on the trilogue's table, for models of 10^24 FLOP**, with great confidence that this would not hinder innovation nor the Brussels Effect for cutting-edge models.

Executive Summary for Policymakers:

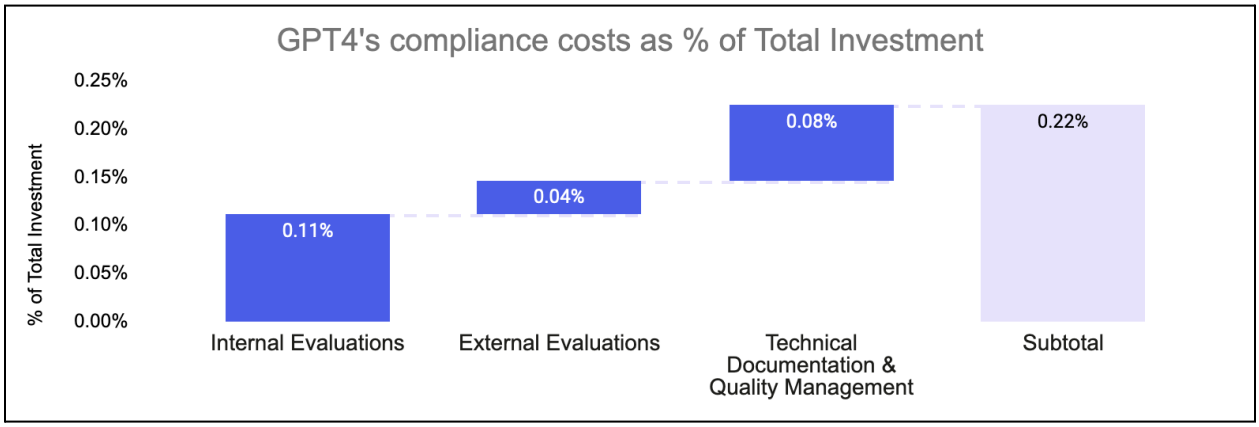
In this memo, we estimate the significance of compliance costs related to the AI Act for General-purpose AI models provider. We do so by estimating the financial effort (i.e. the investment) necessary to develop a cutting-edge model, including the significant expense to secure a significant amount of chips, non-GPU hardware and engineers for long enough to train and deploy such a model. We then estimate the compliance costs involved with the European Parliament's Article 28b on GPAI models. These compliance costs include the internal and external evaluations for managing risks and ensuring appropriate levels of safety and predictability, the technical documentation and the quality management system. We use conservative assumptions throughout (e.g. San Francisco salaries, need for 5 staff full-time for 3 months to ensure compliance of a single model, need for a second external evaluation because of failing the initial one, ...).

We find that compliance costs for GPAI models represent between 0.07% and 1.34% of the total capital expenditure needed for building such a model. Table 1 provides our results and Figure 1 breaks down the costs for GPT4, as an example.

Table 1:

Threshold of training compute (FLOP)	# of Models above that threshold but below the next threshold (private sector)	Number of Providers (private sector)	Total Investment per model (€)	Compliance costs as % of Total Investment per model
10^26	0	0	1,992,524,559	**0.07%**
GPT4 (2.1*10^25)	NA	1	456,513,204	**0.22%**
10^25	2	2 (OpenAI, Inflection (tbc))	234,462,537	**0.42%**
10^24	8	6 (Anthropic, Baidu, Google, Inflection, Microsoft, OpenAI)	60,683,754	**1.34%**

Figure 1:



Our analysis suggests that while the **costs associated with compliance under Article 28b of the European Parliament's AI Act compromise represent a negligible portion of the total investment in developing GPAI models at 10^24 FLOP and beyond.** This expenditure is crucial for ensuring the safety, security, and reliability of these technologies, benefiting EU citizens and the broader digital economy.

Context

As the final trilogue on the EU AI Act approaches, numerous voices across civil society, academia, experts, startups and SMEs have raised concerns about exempting general-purpose AI (GPAI) models. In addition to undermining product safety and generating risks to EU citizens, many of them have pointed out that lack of adequate guarantees for the predictability of a general-purpose AI model could undermine trust in the European single market, and even stifle digital innovation for downstream use cases. Additionally, academics have expressed concerns over the unregulated development and deployment of general-purpose AI models and the risks they pose to fundamental rights. However, alongside these calls for regulation of general-purpose AI models, there remains a degree of uncertainty about the potential costs and burdens associated with regulatory proposals like Article 28b.

This memo offers a preliminary estimate of the additional costs that compliance with Article 28b of the European Parliament compromise text would impose on a general-purpose AI model provider, and how these compliance costs would compare to compute investments. Given the absence of technical standards for Article 28b, the scarcity of data points and the rapidly evolving field of AI, any estimates provided are inherently subject to significant limitations. Nevertheless, by examining current best practices in other industries and in general-purpose AI model development, such as the testing and evaluation procedures of advanced systems like GPT-4, we put forward a rough understanding of the potential costs involved. This is a working document that we will keep updating given future AI development and as we receive new information.

1. Assumptions, Focus and Method

For this analysis, we assume a tiered approach, based on compute thresholds, that would allow to scale requirements for general-purpose AI models in proportion to their risk potential. Since predicting a model's impact, use cases, and scope can be challenging, a model's size, gauged by the cumulative compute power used during training—measured in FLOP (Floating Point Operations)—can serve as a practical ex ante risk indicator. The EU tiered approach's exact thresholds remain hotly debated, with figures ranging from 10^21 to 10^26 FLOP. For our purposes, looking at the bleeding-edge of general-purpose AI model development and given the scarcity of data, we provide estimates for 10^24 to 10^26 FLOP.

We focus on the incremental costs of compliance, particularly in evaluations for risk mitigation, technical documentation and quality management, as required by Article 28b of the European Parliament compromise text of 14/06/2023. We chose this version of the rules on general-purpose AI models because it is the most "heavy-handed" one currently on the negotiating table, placing binding obligations and requirements directly on the shoulders of

general-purpose AI model providers, to be complied with using state-of-the-art methods. This choice ensures our estimates are as conservative as possible. In brief, Article 28b (see Appendix I) mandates various measures throughout the AI lifecycle to manage risks and quality, and ensure performance, predictability, interpretability, corrigibility, safety and cybersecurity.

We believe that using Article 28b as a reference[1] helps us project a fair assessment of the compliance costs, constituting an "upper bound" of how onerous the AI Act's requirements for GPAI models can be. Note that several requirements align with existing best practices at AI companies carried out regardless of the AI Act. For instance, in order to comply with existing laws or to satisfy users' preference, models are already commonly fine-tuned to filter harmful speech, as a way to ensure safety and manage risk. Additionally, some requirements, like using energy-efficient hardware, inadvertently reduce compute and inference costs. Therefore, our focus is on the novel costs uniquely introduced by Article 28b (Appendix I): evaluations (internal & external) to formalize risk management & benchmark-related requirements, technical documentation and the quality management system.

Our sources and methodology are grounded in a combination of publicly available information and expert insights. We used publicly available resources such as Epoch's ML trends database or OpenAI's GPT-4 system card, which provide some indications on e.g., time spent on evaluations and number of risk scenarios analyzed. This was complemented with interviews of experts from leading AI companies and third-party evaluation organizations. Finally, we have also leveraged publicly available information about compliance costs for similar practices in other heavily regulated industries (automotive type approval, medical devices and other machinery-related). Together these sources built the basis to inform our back-of-the-envelope calculations.

2. Total Investment for building GPAI models

Developing a general-purpose AI model such as GPT4 requires a financial effort of ~457 Million €. This is an estimate of the hurdle to develop a cutting-edge model, including the significant expense to secure a significant amount of chips, non-GPU hardware and engineers for long enough to train and run such a model. It considers GPT4 as an investment rather than a product, given the increasing role of these GPAI models as platforms or assets that can serve multiple instances, which can be plugged in various applications or fine tuned to become any application, as opposed to a mere product.[2] The breakdown in capital expenditure to build that asset is provided on Table 2 below. This finding is consistent with the empirical evidence that there are few AI companies that can afford the capital expenditure required to build such models, or even models with lower amount of compute (6 companies for models above 10^24 FLOP), and

[1] Rather than the far less constraining proposals currently discussed, such as approaches relying on self-regulation or on light-touch risk mitigation, or on unreasonably high thresholds.
[2] Cf. Appendix II - Footnote 2.

therefore few such models have ever been built worldwide (10 models above 10^24 FLOP) (cf. Table 1).

Table 2: Estimates of cost factors (all figures in €)

Threshold of training compute (FLOP)	Total capital expenditure (excl. compliance cost)	GPU costs	Other hardware costs	Labor costs for 1 year	Energy costs
10^26	**1,992,524,559**	1,359,710,000	582,732,856	41,375,000	8,706,703
GPT4 (2.15*10^25)	**456,513,204**	289,300,000[3]	123,985,714[4]	41,375,000[5]	1,852,490
10^25	**234,462,537**	134,558,140	57,667,774	41,375,000	861,623
10^24	**60,683,754**	13,455,814	5,766,777	41,375,000	86,162

Note that we assume GPU costs, other hardware costs and energy costs scale linearly with the training compute per model, while labor costs remain constant across all thresholds.

3. Compliance Costs Under Article 28b for GPAI models

Complying with Article 28b is estimated to cost 0.07% and 1.34% of the Total Investment for general-purpose AI models. Article 28b would require a GPAI model developer to, for instance, ensure safety and performance through "appropriate methods such as model evaluation with the involvement of independent experts, documented analysis, and extensive testing during conceptualisation, design, and development" as well as technical documentation and quality management systems (EP, p198; see Appendix I). Below, we estimate additional costs that Article 28b would introduce, i.e. internal evaluations, external evaluation as well as technical documentation and quality management systems, to achieve the risk management and appropriate levels of safety and predictability.

[3] Cf. Appendix II - Footnote 3.

[4] Cf. Appendix II - Footnote 4.

[5] Personnel Cost for GPT-4 Training: The estimated personnel cost for GPT-4's development is derived from OpenAI's reported expenses for 2022. Out of their total expenditure, $89.31 million was specifically attributed to staff costs. We divided this figure by two, estimating conservatively that only half of OpenAI staff were working on developing and training GPT-4.

Table 3: Compliance costs as percentage of total investment per model

Threshold of training compute (FLOP)	Total compliance costs	Internal evaluations	External evaluations	Technical documentation & Quality Management system
10^26	0.07%	0.03%	0.02%	0.02%
GPT4	0.22%	0.11%	0.04%	0.08%
10^25	0.42%	0.21%	0.06%	0.15%
10^24	1.34%	0.59%	0.16%	0.59%

Internal Evaluations: 362,500€ or 0.59% of Total Investment for a 10^24 FLOP GPAI model. The estimated cost for companies to conduct internal evaluations is calculated based on salaries of in-house employees. Per model, we conservatively assume an additional internal evaluation team of 5 full-time equivalent (FTE) for 3 months with annual San Francisco-based annual salaries ranging 200,000€ (3 FTE operations and legal officers for 3 months) to 424,906€[6] (2 FTE Professional Red Teamers/Research Scientists for 3 months). This sums up to approximately 362,500€. Note that this 5-FTE team is already in addition to the red-teaming system that all of the companies affected have already established. Moreover, San Francisco-based salaries are usually 2 to 3 times higher than EU equivalent.

Increasingly capable models are more problematic in terms of alignment, predictability and overall safety, as has been documented empirically.[7] As a result, we assume that each order of magnitude in additional compute (which is a proxy for generality of capabilities) increases the duration of internal evaluation by one third.[8] The results are in table 3. These represent between 0.59% and 0.03% of total investment per model (362,500 to 644,042€).

External Evaluations: 101,162€ or 0.16% of Total Investment for a 10^24 FLOP GPAI model. Currently, external evaluations, such as performance benchmarking via the HELM benchmark or dangerous capabilities evaluations are performed by academia or non-profits, such as METR

[6] Based on the job posting salary for a Research Scientist, Frontier Red Team (Cyber) at Anthropic, the highest paid we found.
[7] [2212.09251] Discovering Language Model Behaviors with Model-Written Evaluations
[8] 10^24 = 3 months; 10^25 = 4 months; GPT4 = 4.20 months; 10^26 = 5.33 months; which is consistent with observed practice. Note that this computation conservatively refers to the additional red teaming & evaluations needed for 28b specifically, not the red teaming and evaluation carried out already by providers.

(formerly ARC). This means that external evaluation costs are not representative of actual market costs. A mature market for Article 28b's external evaluations would converge on much more systematic procedures, akin to product or quality audits. After discussing this cost with experts and using costs in regulatorily comparable industries, we've estimated that 101,162€ per 10^24 model would be needed for external evaluations. These costs do not scale linearly with models' compute training size, as some of the costs of an external evaluation are fixed and some are dependent on being at the cutting edge: 10^25 would be 137,948€ (0.06%), GPT4 would be 160,939€ (0.04%), and 10^26 would be 410,201€ (0.02%). Our assumptions are in Appendix II.

Technical Documentation and Quality Management Systems: 360,000€ or 0.59% of Total Investment for a 10^24 FLOP GPAI model. The 2021 Impact Assessment of the European Commission for the AI Act found that technical documentation would amount to roughly 20,000€ - 30,000€ for a down-stream high-risk AI system.[9] A quality management system would add to roughly €330,000. Given the generality of general-purpose AI models and potential risk scenarios, this figure would tend to be somewhat higher; however, Article 28b technical documentation and quality management requirements are significantly less stringent than for high-risk AI systems, and leave room for experimentation - offsetting the greater product complexity. There is no variation with various compute thresholds. In total, we therefore find a flat 360,000€ cost (0.59% of Total Investment for a 10^24 FLOP and 0.02% for a 10^26 FLOP model).

[9] Impact Assessment of the Regulation on Artificial intelligence | Shaping Europe's digital future

Appendices

Appendix I - Article 28b

Article 28b
(European Parliament, p 198); Excerpt:

Obligations of the provider of a foundation model

1. A provider of a foundation model shall, prior to making it available on the market or putting it into service, ensure that it is compliant with the requirements set out in this Article, regardless of whether it is provided as a standalone model or embedded in an AI system or a product, or provided under free and open source licences, as a service, as well as other distribution channels.

2. For the purpose of paragraph 1, the provider of a foundation model shall:

> *(a) demonstrate through appropriate design, testing and analysis the identification, the reduction and mitigation of reasonably foreseeable risks to health, safety, fundamental rights, the environment and democracy and the rule of law prior and throughout development with appropriate methods such as with the involvement of independent experts, as well as the documentation of remaining non-mitigable risks after development*
>
> *(b) process and incorporate only datasets that are subject to appropriate data governance measures for foundation models, in particular measures to examine the suitability of the data sources and possible biases and appropriate mitigation*
>
> *(c) design and develop the foundation model in order to achieve throughout its lifecycle appropriate levels of performance, predictability, interpretability, corrigibility, safety and cybersecurity assessed through appropriate methods such as model evaluation with the involvement of independent experts, documented analysis, and extensive testing during conceptualisation, design, and development;*
>
> *(d) design and develop the foundation model, making use of applicable standards to reduce energy use, resource use and waste, as well as to increase energy efficiency, and the overall efficiency of the system, whithout [sic] prejudice to relevant existing Union and national law. This obligation shall not apply before the standards referred to in Article 40 are published. Foundation models shall be designed with capabilities enabling the measurement and logging of the consumption of energy and resources, and, where technically feasible, other environmental impact the deployment and use of the systems may have over their entire lifecycle;*
>
> *(e) draw up extensive technical documentation and intelligible instructions for use, in order to enable the downstream providers to comply with their obligations pursuant to Articles 16 and 28(1);.*
>
> *(f) establish a quality management system to ensure and document compliance with this Article, with the possibility to experiment in fulfilling this requirement,*

(g) register that foundation model in the EU database referred to in Article 60, in accordance with the instructions outlined in Annex VIII point C.

When fulfilling those requirements, the generally acknowledged state of the art shall be taken into account, including as reflected in relevant harmonised standards or common specifications, as well as the latest assessment and measurement methods, reflected in particular in benchmarking guidance and capabilities referred to in Article 58a;

3. Providers of foundation models shall, for a period ending 10 years after their foundation models have been placed on the market or put into service, keep the technical documentation referred to in paragraph 2(e) at the disposal of the national competent authorities

4. Providers of foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("generative AI") and providers who specialise a foundation model into a generative AI system, shall in addition

a) comply with the transparency obligations outlined in Article 52 (1),
b) train, and where applicable, design and develop the foundation model in such a way as to ensure adequate safeguards against the generation of content in breach of Union law in line with the generally-acknowledged state of the art, and without prejudice to fundamental rights, including the freedom of expression,
c) without prejudice to Union or national or Union legislation on copyright, document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law.

Appendix II – Calculations & Assumptions

Sources for Total Investment per model:

Footnote 2: We utilize the concept of 'investment' rather than 'costs' for several reasons:

- **Resource Utilization Relative to Investment:** The burden of compliance should be measured in terms of the resources expended relative to the resources invested in developing the model; rather than the amortized accounting cost (approximated by cloud rental), the capital is the bottleneck.
- **Technological Frontier Constraints:** At the cutting-edge of technology such as 10^24 FLOP models, cloud service providers often do not possess sufficiently large GPU clusters to meet the one-off demand of training of advanced models (though cloud providers definitely are key for model serving at scale). Consequently, model developers are often compelled to invest in their own compute chips.

- **Trends in Hardware Acquisition:** Experts noted that leading-edge providers frequently acquire their own semiconductor chips, reinforcing the need for a more investment-centric view. Notably, OpenAI, despite its partnership with cloud giant Microsoft Azure, has developed and aims to maintain its own chip cluster.
- **Compliance Costs as an Investment:** Compliance costs, which we estimate later in this study, should be viewed as an investment in the organization's assets. Beyond the quality of the asset produced, compliance enhances the company's brand reputation, access to business-to-business markets, among others. Therefore, it is appropriate to compare one form of investment (compliance costs) with another (compute investment). In our analysis, we disregard the residual value of compliance, paralleling our omission of the residual value of hardware components such as chips.

Footnote 3:
The estimated GPU cost for training GPT-4 is about €289.3 million, based on two distinct estimation methods.

1. GPU-Specific Estimate: Based on the A100 accelerator unit costs, we estimate a requirement for between 20,000 and 30,000 units, averaging at 25,000 units. With each unit priced between $10,000 and $15,000, and taking an average cost of $12,500 per unit, the total cost ranges up to approximately $312.5 million, which is €289,078, 931. For the A100 accelerator units, required for GPT-4's training, an estimated 20,000 to 30,000 units were needed, with an average of 25,000 units considered. Consequently, the total expenditure is projected to be around $312.5 million, equivalent to approximately €289.3 million.
2. Expenditure-Based Estimate: Derived from OpenAI's reported 2022 expenditure of over $400 million on computing and data, assuming 70% allocated to GPUs, in line with industry norms for ML supercomputers. This gives $280MM, equivalent to approximately €259,3MM.

Footnote 4:
Other Hardware Cost for GPT-4 Training: In addition to the GPU costs, the remaining hardware expenses for GPT-4 are estimated at €111.1 million, mostly networking hardware to allow fast communication between chips - dataset building costs are negligible at such a scale. This figure is calculated as 30% of the total compute cost, which is determined to be approximately €370.43 million. This total compute cost is deduced based on the known GPU expenditure (€289.3 million) forming 70% of the overall compute budget, in line with industry norms for ML supercomputers.

Calculations for Compliance costs:

External Evaluations:
Our assumptions are that the daily rate of external evaluator/independent researchers/notified body's experts is $3,000.

Number of expert-days needed: 10 days per month of evaluation procedure.

External evaluation procedure will require 1 month (for easily replicable evals) or up to 3 months (for stringent, state-of-the-art evals).

The GPAI model provider dedicates 1.5 FTEs to support external evaluators, paid San Francisco-based annual salaries ranging 200,000€ (1 cumulative FTE of operations/legal officers) and 424,906€[10] (0.5 FTE Professional Red Teamers/Research Scientists).

A failed evaluation will require a second evaluation. We expect this to occur 10% of the time for a 10^24 model, followed by a successful second evaluation, leading to an expected 1.10 external evaluation needed per model. This likelihood increases progressively towards 90% for 10^26 models, leading to an expected 1.9 external evaluation.[11] Note that this evaluation is to verify the compliance of the model, and is different from existing "red-teaming" and "private beta testing", which are carried out multiple times already regardless of the AI Act; and which do not constitute an official assessment of compliance.

We further conservatively assume that models at the cutting edge (10^26 FLOP) will require a stringent evaluation as opposed to normal evaluation, to reflect the unknown unknowns and additional scrutiny from developing such poorly understood yet powerful models.

The external evaluation provider's flat fees & non-expert wages, to cover all support costs beyond expert-days are estimated to be 30,000€. This is based on estimates of compliance and conformity assessment costs across various industries: 4,500€ (safety of low-voltage equipment); 8500€ (automotive type approval), 10,000€ (class II Medical devices), 50,000€ (class III Medical devices), 55,000€ (construction products), consistent with the range 100€ to 100,000€ almost unanimously quoted by practitioners. None of the costs included in this category (drafting technical documents, declaration/certification, controls/audits, wage for compliance controller/auditor) are expected to scale with the model's power.

[10] Based on job posting salary for a Research Scientist, Frontier Red Team (Cyber) at Anthropic, the highest paid we found.

[11] 50% for 10^25, and 75% for GPT4. Note that we assume that the second evaluation is successful 100% of the time for any model; if it weren't, it would mean that we are dealing with a severely problematic model/provider who likely does not invest enough in compliance to pass the external evaluation, and their "compliance costs" would not be representative.

THE FUTURE SOCIETY

867 Boylston Street
Boston, MA 02166