stability.ai

The AI Act should promote grassroots innovation in open models

Background

The AI ecosystem is diverse, from corporate labs building closed-source products to everyday developers sharing open technology. In this environment, open models play a vital role in helping to promote transparency and competition in AI.[1] Today, open models are driving a wave of grassroots innovation among developers, researchers, and small businesses across Europe:

- **Open models promote transparency.** Researchers and authorities can "look under the hood" of an open model to verify performance, identify risks or vulnerabilities, study interpretability techniques, and implement new mitigations. By comparison, closed models may not disclose how they are developed or how they operate. Closed models may be comparatively opaque, and risk management may depend on trust in the developer.

- **Open models lower barriers to entry.** Training a new "base" model from scratch requires significant resources that are not available to everyday developers.[2] Open models lower these barriers to entry. European developers can build on open models to create new AI tools or launch new AI ventures without spending tens of millions of euros on research and computing.

- **Open models drive innovation in safety.** Developers can customize open models for improved safety or performance in specific tasks. For example, open models can be optimized through a range of techniques to mitigate undesirable behavior such as bias, misinformation, or toxicity (e.g. via fine-tuning or reinforcement learning). These techniques can yield significant improvements in the performance of a model without requiring extensive research or computing. That means ordinary developers can build safer and more effective models to better support real-world applications.[3]

- **Open models foster strategic independence.** AI models will be essential infrastructure across the digital economy. They will transform access to services, reshape how we

[1] Open models are software programs that are released publicly along with the billions of distinctive settings or "parameters" that determine the model's performance.
[2] OpenAI disclosed that it cost USD 100 million to train the closed-source GPT-4 model: Wired, 'Open AI's CEO says the age of giant models is already over', April 2023, available here.
[3] Hugging Face, an AI repository, tracks the evaluation results for over 1200 open models on the 'Open LLM Leaderboard', available here.

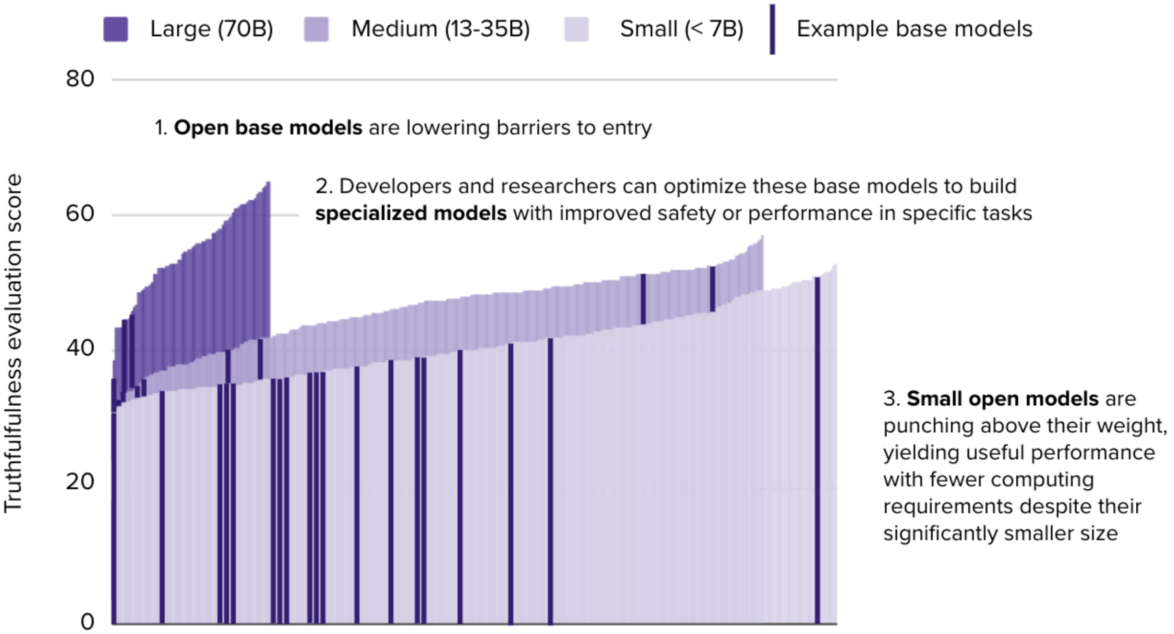
search and access information, and support knowledge management and decision making in some of our most important public and private institutions. Open models enable organizations to build sovereign AI capabilities without relying on a handful of firms for foundational technology. They can develop these AI capabilities “in house” without exposing their data or ceding control of their AI models to foreign firms.

- **Open models help to make AI accessible.** Smaller open models are helping to make AI more efficient, more accessible, and more useful. Unlike large models, which require significant computational resources, small models can deliver useful performance with regular hardware. These models may be hundreds of times smaller than a large closed-source model such as GPT-4. Users can run small models on local devices, including smartphones, and developers can train these models with desktop hardware.

Open models promote fair access to foundational technology. They put these capabilities in the hands of the frontline developers, researchers, and organizations who can best decide how they should be used. They enable developers to build AI systems that are safe, secure, and fit for purpose. And they help to make AI more accessible and more useful. The European Union should actively nurture this grassroots innovation in open models.

Open models drive grassroots innovation in AI safety and performance

Evaluation scores for “truthful” behavior in ~760 open language models



1. **Open base models** are lowering barriers to entry

2. Developers and researchers can optimize these base models to build **specialized models** with improved safety or performance in specific tasks

3. **Small open models** are punching above their weight, yielding useful performance with fewer computing requirements despite their significantly smaller size

Source: TruthfulQA evaluation scores on the Hugging Face Open LLM Leaderboard (September 2023). Models sorted by ascending score. TruthfulQA measures a model’s tendency to reproduce falsehoods (a higher score is better). Model size in billions of parameters.

Challenge

Stability AI has welcomed our engagement with EU institutions over the summer, and we applaud the EU's stated commitment to support open innovation in Europe. For example, the European Parliament's draft text recognizes the importance of open-source AI components, and aims to exempt these from unnecessarily stringent regulation (see Recitals 12a-c).

However, as drafted, the Act may unintentionally stifle grassroots innovation in open models. The binding text of the Parliamentary draft excludes models from the open-source exemption. Instead, it adopts a "one size fits all" approach to regulation. The binding text will treat all models identically, regardless of whether they are powerful new "base" models, or simply "fine-tuned" performance improvements to an existing model, and regardless of whether they are highly versatile models, or models with narrow capabilities. Further, the binding text will capture all kinds of public releases, regardless of whether they are models released in the course of a commercial activity or AI system deployment, or simply models released for information sharing and collaborative research.

The proposed obligations for models may be feasible for a sophisticated corporate developer with a well-resourced compliance department. However, they are unlikely to be feasible for the everyday developers and independent researchers who share, use, or contribute to thousands of open models today. As drafted, the Act is likely to have a significant chilling effect on collaborative research and grassroots innovation. That would represent a major setback for AI safety, and may jeopardize EU competitiveness in AI development and deployment.

Path forward

The Act should adopt a risk-based approach to model regulation – as it does for AI systems – with obligations that are proportional to how a model is used or intended to be used. The Act should promote the sharing of free and open models by everyday developers and independent researchers in Europe, while ensuring that models released in the course of a commercial activity or deployed in AI systems of concern are subject to robust oversight. Model development should be regulated cautiously to avoid stifling collaborative research and grassroots innovation. To that end, we encourage the following modest proposals to avoid unintended consequences:

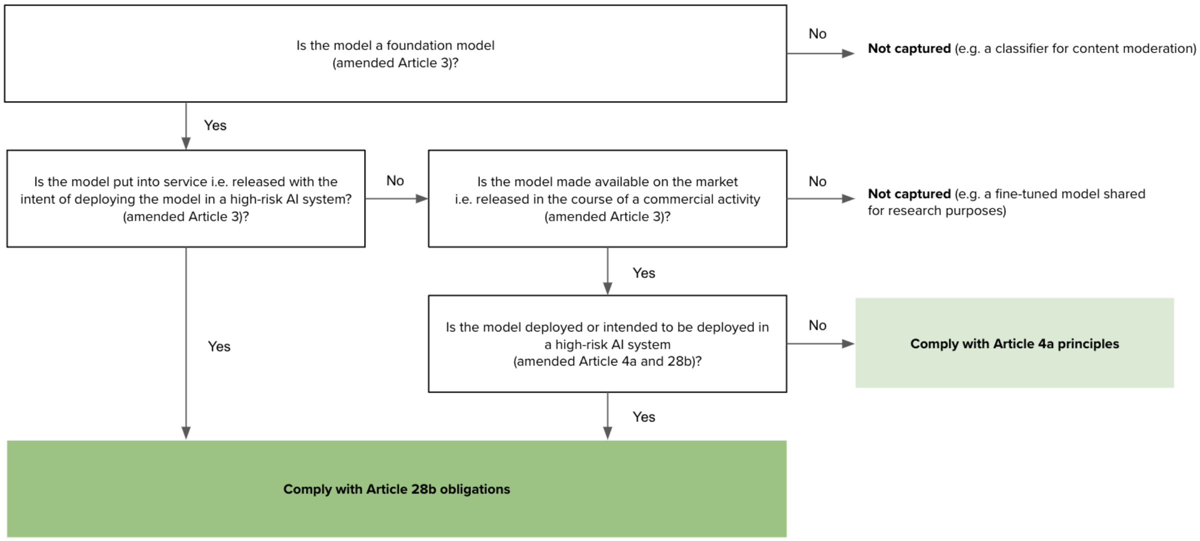
Step 1. **Is the model a regulated model (Article 3)?** The proposed changes clarify that only models that meet certain criteria for capability and versatility will be captured by the Act. There are a range of narrow AI models that should not be considered "foundation models" for the purposes of the Act, such as analytic models (e.g. for modeling weather) or classification models (e.g. for content moderation), but are caught by the draft language.

Step 2. **Is the model available on the market or put into service (Article 2 and Article 3).** The recitals imply that free and open models released on a repository will not be considered on the market or in service. However, this interpretation is not reflected in the binding text, which stipulates that open-source exemptions do not apply to models (Article 2), and that "putting into service" includes any provision of a model to a deployer (Article 3). The proposed changes clarify that free and open models will not be considered "on the market" unless in the course of a commercial activity and will not be considered "put into service” unless they are intended to be deployed in an AI system of concern (e.g. a high-risk AI system).

Step 3. **Is the model used or intended to be used in an AI system of concern (Article 4a and Article 28b)?** The proposed changes emphasize that models that are deployed or intended to be deployed in AI systems of concern must comply with the mandatory obligations in Article 28b. Further, the proposed changes stipulate that high-risk AI systems can *only* use a model that complies with Article 28b. However, open models that are made available on the market or put into service for other purposes (e.g. research, information sharing, or low-risk applications) need only comply with the baseline principles outlined in Article 4a.

Together, these proposals adopt a risk-based approach for the regulation of models. They ensure that Article 28b obligations apply to open models that are deployed or intended to be deployed in a high-risk AI system; they retain Article 4a obligations for open models that are made available on the market for lower-risk AI deployments; and they avoid burdening open models released for other purposes, such as information sharing or collaborative research. These modest amendments can help to sustain grassroots innovation in Europe.

Effect of our proposed amendments for open models


Is the model a foundation model
(amended Article 3)?
No
Not captured (e.g. a classifier for content moderation)
Yes
Is the model put into service i.e. released with the intent of deploying the model in a high-risk AI system? (amended Article 3)?
No
Is the model made available on the market i.e. released in the course of a commercial activity (amended Article 3)?
No
Not captured (e.g. a fine-tuned model shared for research purposes)
Yes
Yes
Is the model deployed or intended to be deployed in a high-risk AI system (amended Article 4a and 28b)?
No
Comply with Article 4a principles
Yes
Comply with Article 28b obligations

Suggested amendments
*Amendments in **bold***

Parliament text	Amended text	Explanation
Art 2 – para 5d *This Regulation shall not apply to research, testing and development activities regarding an AI system prior to this system being placed on the market or put into service, provided that these activities are conducted respecting fundamental rights and the applicable Union law. The testing in real world conditions shall not be covered by this exemption. The Commission is empowered to adopt delegated acts in accordance with Article 73 that clarify the application of this paragraph to specify this exemption to prevent its existing and potential abuse. The AI Office shall provide guidance on the governance of research and development pursuant to Article 56, also aiming to coordinate its application by the national supervisory authorities;*	*Art 2 – para 5d* *This Regulation shall not apply to research, testing and development activities regarding an AI system **or foundation model** prior to this system being placed on the market or put into service **within the meaning of Article 3 [as amended below]**, provided that these activities are conducted respecting fundamental rights and the applicable Union law. The testing in real world conditions shall not be covered by this exemption. The Commission is empowered to adopt delegated acts in accordance with Article 73 that clarify the application of this paragraph to specify this exemption to prevent its existing and potential abuse. The AI Office shall provide guidance on the governance of research and development pursuant to Article 56, also aiming to coordinate its application by the national supervisory authorities;*	This amendment clarifies that research and development exemptions also apply to foundation models, prior to being made available on the market or put into service within the meaning of Article 3 as amended (below).
Article 2 – para 5e *This Regulation shall not apply to AI components provided under free and open-source licences except to the extent they are placed on the market or put into service by a provider as part*	*Article 2 – para 5e* *This Regulation shall not apply to AI components provided under free and open-source licences except to the extent they are placed on the market or put into service by a provider as part*	The existing language ("shall not apply…") implies that free and open foundation models will be subject to the Act regardless of whether or not they have been placed on the market or put into service. However, free and open

of a high-risk AI system or of an AI system that falls under Title II or IV. This exemption shall not apply to foundation models as defined in Art 3.	*of a high-risk AI system or of an AI system that falls under Title II or IV. This exemption shall not apply to foundation models* ***that are made available on the market in the course of a commercial activity or put into service*** *as defined in Art 3.*	foundation models that have not been placed on the market or put into service within the meaning of Article 3 (as clarified below) should not be subject to the Act, consistent with Recital 12b: "Neither the collaborative development of free and open-source AI components nor making them available on open repositories should constitute a placing on the market or putting into service."
Article 3 – para 1 – point (1c) *'foundation model' means an AI system model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks;*	*Article 3 – para 1 – point (1c)* *'foundation model' means ~~an~~ AI system model* ***software that is (i) intended to process inputs from an AI system and return outputs, (ii) is designed for a broad range of applications as determined by [the implementing authority], (iii) the performance of which is determined predominantly by automated training on large datasets rather than pre-programmed rules,*** *~~an AI system model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks,~~* ***and (iv) meets other criteria for capability determined by [the implementing authority];***	The definition of "foundation model" is vague and overbroad, making it difficult for grassroots developers to anticipate whether the Act will apply to their activities. Further, the existing definition captures a range of AI models that should not be considered foundation models, such as classifiers to detect or moderate content, or analytic models to support business forecasts and optimization. The proposed changes align with the commitment in Recital 60g: "pre-trained models developed for a narrower, less general, more limited set of applications… should not be considered foundation models for the purposes of this Regulation".
Article 3 – para 1 – point 10 *'making available on the market' means any supply of an AI system for distribution or use on the Union market in*	*Article 3 – para 1 – point 10* *'making available on the market' means any supply of an AI system for distribution or use on the Union market in*	Open models help to promote transparency in AI; improve access to critical technology; and support the development of safer, fairer, and more

the course of a commercial activity, whether in return for payment or free of charge; *Article 3 – para 1 – point 11* *'putting into service' means the supply of an AI system for first use directly to the deployer or for own use on the Union market for its intended purpose;*	*the course of a commercial activity, whether in return for payment or free of charge****, but does not include the free and open sharing of a component or model unless in the course of a commercial activity****;* *Article 3 – para 1 – point 11* *'putting into service' means the supply of an AI system for first use directly to the deployer or for own use on the Union market for its intended purpose,* ***but does not include the free and open sharing of a component or model unless it is intended to be deployed in a high-risk AI system;***	effective models. The Act should continue to promote grassroots innovation in open models. However, as drafted, these provisions do not clearly exempt the sharing of free and open models. These provisions do not give effect to Recital 12b: "Neither the collaborative development of free and open-source AI components nor making them available on open repositories should constitute a placing on the market or putting into service." Releasing a model on a free and open basis should not constitute "making available on the market" unless in the course of a commercial activity. Likewise, it should not constitute "putting into service" unless the model is intended to be deployed in an AI system of concern, such as a high-risk AI system. For example, an independent researcher sharing optimized or "fine-tuned" models should not be subject to the same requirements as a sophisticated corporate actor releasing a powerful "base" model for the first time. When applying these provisions, intent may be inferred objectively from the circumstances of the release, including any representations made by the model developer (e.g. about the suitability of

		the model); relationships between the model developer and downstream AI system providers or deployers (e.g. contracts or partnerships); and the direct or indirect interests of the model developer in releasing the model (e.g. present or future financial interests).
Article 4a – para 1 *All operators falling under this Regulation shall make their best efforts to develop and use AI systems or foundation models in accordance with the following general principles establishing a high-level framework that promotes a coherent human-centric European approach to ethical and trustworthy Artificial Intelligence, which is fully in line with the Charter as well as the values on which the Union is founded:* *a) 'human agency and oversight' means that AI systems shall be developed and used as a tool that serves people, respects human dignity and personal autonomy, and that is functioning in a way that can be appropriately controlled and overseen by humans;* *b) 'technical robustness and safety' means that AI systems shall be developed and used in a way to minimize unintended and unexpected harm as well as being robust in case of unintended problems and being resilient*	*Article 4a – para 1* *All operators falling under this Regulation shall make their best efforts to develop and use AI systems or foundation models in accordance with the following general principles establishing a high-level framework that promotes a coherent human-centric European approach to ethical and trustworthy Artificial Intelligence, which is fully in line with the Charter as well as the values on which the Union is founded:* *a) 'human agency and oversight' means that AI systems shall be developed and used as a tool that serves people, respects human dignity and personal autonomy, and that is functioning in a way that can be appropriately controlled and overseen by humans;* *b) 'technical robustness and safety' means that AI systems shall be developed and used in a way to minimize unintended and unexpected harm as well as being robust in case of unintended problems and being resilient*	The proposed amendments clarify that if a model is placed on the market or put into service (as defined above), it must comply with Article 4a design principles at a minimum. The proposed amendments stipulate that additional Article 28b requirements will apply if the model is deployed or intended to be deployed in an AI system of concern (e.g. a high-risk AI system), without imposing those requirements on open models that are made available on the market or put into service for low-risk applications.

against attempts to alter the use or performance of the AI system so as to allow unlawful use by malicious third parties; *c) ‘privacy and data governance’ means that AI systems shall be developed and used in compliance with existing privacy and data protection rules, while processing data that meets high standards in terms of quality and integrity;* *d) ‘transparency’ means that AI systems shall be developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system as well as duly informing users of the capabilities and limitations of that AI system and affected persons about their rights;.* *e) ‘diversity, non-discrimination and fairness’ means that AI systems shall be developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law;* *f) ‘social and environmental well-being’ means that AI systems shall be developed and used in a sustainable and environmentally friendly*	*against attempts to alter the use or performance of the AI system so as to allow unlawful use by malicious third parties;* *c) ‘privacy and data governance’ means that AI systems shall be developed and used in compliance with existing privacy and data protection rules, while processing data that meets high standards in terms of quality and integrity;* *d) ‘transparency’ means that AI systems shall be developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system as well as duly informing users of the capabilities and limitations of that AI system and affected persons about their rights;.* *e) ‘diversity, non-discrimination and fairness’ means that AI systems shall be developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law;* *f) ‘social and environmental well-being’ means that AI systems shall be developed and used in a sustainable and environmentally friendly*	

manner as well as in a way to benefit all human beings, while monitoring and assessing the long-term impacts on the individual, society and democracy. *2. Paragraph 1 is without prejudice to obligations set up by existing Union and national law. For high-risk AI systems, the general principles are translated into and complied with by providers or deployers by means of the requirements set out in Articles 8 to 15, and the relevant obligations laid down in Chapter 3 of Title III of this Regulation. For foundation models, the general principles are translated into and complied with by providers by means of the requirements set out in Articles 28 to 28b. For all AI systems, the application of the principles referred to in paragraph 1 can be achieved, as applicable, through the provisions of Article 28, Article 52, or the application of harmonised standards, technical specifications, and codes of conduct as referred to in Article 69, without creating new obligations under this Regulation.*	*manner as well as in a way to benefit all human beings, while monitoring and assessing the long-term impacts on the individual, society and democracy.* *2. Paragraph 1 is without prejudice to obligations set up by existing Union and national law. For high-risk AI systems, the general principles are translated into and complied with by providers or deployers by means of the requirements set out in Articles 8 to 15, and the relevant obligations laid down in Chapter 3 of Title III of this Regulation. For foundation models, the general principles* ***apply to all foundation models that are made available on the market or put into service*** ~~*are translated into and complied with by providers by means of the requirements*~~ ***with additional requirements*** *set out in Articles 28 to 28b* ***for models that are deployed or intended to be deployed in high-risk AI systems****.* For *all AI systems, the application of the principles referred to in paragraph 1 can be achieved, as applicable, through the provisions of Article 28, Article 52, or the application of harmonised standards, technical specifications, and codes of conduct as referred to in Article 69, without creating new obligations under this Regulation.*	

Article 28 – para 2 *Where the circumstances referred to in paragraph 1, point (a) to (ba) occur, the provider that initially placed the AI system on the market or put it into service shall no longer be considered a provider of that specific AI system for the purposes of this Regulation. This former provider shall provide the new provider with the technical documentation and all other relevant and reasonably expected information capabilities of the AI system, technical access or other assistance based on the generally acknowledged state of the art that are required for the fulfilment of the obligations set out in this Regulation.* *This paragraph shall also apply to providers of foundation models as defined in Article 3 when the foundation model is directly integrated in an high-risk AI system.*	*Article 28 – para 2* *Where the circumstances referred to in paragraph 1, point (a) to (ba) occur, the provider that initially placed the AI system on the market or put it into service shall no longer be considered a provider of that specific AI system for the purposes of this Regulation. This former provider shall provide the new provider with the technical documentation and all other relevant and reasonably expected information capabilities of the AI system, technical access or other assistance based on the generally acknowledged state of the art that are required for the fulfilment of the obligations set out in this Regulation.* *~~This paragraph shall also apply to providers of foundation models as defined in Article 3 when the~~* ***When a*** *foundation model,* ***as defined in Article 3,*** *is directly integrated in an high-risk AI system,* ***the provider of the high-risk AI system will obtain all necessary documentation, information, capabilities, and access from the foundation model provider prior to making available or putting into service the high-risk AI system.***	If the provider of a high-risk AI system integrates a foundation model – especially a free and open model – the burden should fall on the AI system provider to obtain the necessary documentation, information, or access prior to releasing the high-risk AI system. The Act should not compel a developer of free and open models to offer support to downstream providers of AI systems if they neither control how the provider uses the model nor benefit from the AI system. As drafted, these obligations would stifle open development and open innovation in Europe.
Art 28b – para 1 *A provider of a foundation model shall, prior to making it*	*Art 28b – para 1* *A provider of a foundation model* ***intended to be used in***	As drafted, Article 28b does not adopt a risk-based approach to regulation. Article 28b treats grassroots

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.	***a high-risk AI system*** *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.* ***A high-risk AI system that is subject to this Act must use a foundation model that is compliant with this Article.***	developers or independent researchers the same as corporate firms; treats all models identically regardless of capability or adaptability; and avoids the risk-based and proportionate approach that applies to AI systems. These provisions are likely to stifle grassroots development in Europe, set back open innovation, and limit AI development to a handful of technology firms. Instead, the Act should limit the application of Article 28b to models that are intended to be used in an AI system of concern (e.g. a high-risk AI system). The Act should require that high-risk AI systems must use a model that is compliant with Article 28b. Other models will be subject to the "best efforts" requirements in Article 4a, such as open models released for information sharing or collaborative research. In determining the applicability of Article 28b, intent may be inferred objectively from the circumstances of the release, including the representations made by the model developer; relationships between the model developer and an AI system provider or deployer; and the direct or indirect interests of the model developer in releasing the model.

Art 28b – para 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*	*Art 28b – para 2* *For the purpose of paragraph 1, the provider of a foundation model* ***that is subject to this Article*** *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*	Grassroots developers and independent researchers cannot feasibly comply with the requirements of Article 28b(2)-(4). These provisions are likely to stifle grassroots development of open models in Europe. Everyday developers and independent researchers: • Are unlikely to have the resources to comply with the significant procedural requirements in paragraph (2) points (c), (e), (f), or (g); paragraph (3); or paragraph (4)(c); • May not have access to the original data necessary to ensure compliance with dataset obligations in (b), particularly if they are fine-tuning or customizing an existing model; • May not have access to “state of the art” methods or techniques, which are likely to be determined by a small number of well-resourced AI firms. Instead, free and open models released by grassroots developers for purposes other than deployment in a regulated AI system should be excluded from Article 28b (see

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, without 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*	*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, without 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*	amendments to Article 28b(1) above). These models should be subject to “best efforts” obligations for oversight, robustness, safety, data governance, transparency, fairness, and social and environmental wellbeing in Article 4a. Further, compliance with Article 28b(2) should be determined based on “widely-available” methods and “best practices”, not on proprietary “state of the art” techniques pioneered by a small handful of firms. Terminology such as “widely-available” and “best practices” will help to ensure that compliance is feasible for all developers.

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; Art 28b – para 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 Art 28b – para 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	accordance with the instructions outlined in Annex VIII point C. When fulfilling those requirements, the generally acknowledged ~~state of the art~~ **best practices** shall be taken into account, including as reflected in relevant harmonised standards or common specifications, as well as the latest **widely-available** assessment and measurement methods, reflected in particular in benchmarking guidance and capabilities referred to in Article 58a; Art 28b – para 3 Providers of foundation models **that are subject to this Article** 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 Art 28b – para 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	

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.*	*outlined in Article 52 (1)* ***where applicable****,* *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.*	
Art 52 – para 1 *Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that the AI system, the provider itself or the user informs the natural person exposed to an AI system that they are interacting with an AI system in a timely, clear and intelligible manner, unless this is obvious from the circumstances and the context of use.* *Where appropriate and relevant, this information shall also include which functions are AI enabled, if there is human oversight, and who is*	**See amendments to Article 28b(4)(a).**	These disclosure requirements apply to the AI systems that interact with users. It is unclear how a foundation model can comply with these requirements as a standalone component. Article 28b(4)(a) should be amended (above) to clarify that Article 52(1) applies to foundation model providers only “where applicable”.

responsible for the decision-making process, as well as the existing rights and processes that, according to Union and national law, allow natural persons or their representatives to object against the application of such systems to them and to seek judicial redress against decisions taken by or harm caused by AI systems, including their right to seek an explanation. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.		
Article 53 – para 1d *AI regulatory sandboxes shall, in accordance with criteria set out in Article 53a, provide for a controlled environment that fosters innovation and facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan agreed between the prospective providers and the establishing authority.*	N/A – comment only	Sharing models openly is essential to promote research and development into performance, risk, and interpretability. However, the sandboxes proposed by the Act do not support grassroots developers or independent researchers who release free and open models for purposes other than deployment in a regulated AI system. By definition, sandboxing in a “controlled environment” would restrict access to these models, stifling the public exchange of ideas, knowledge, or research that is essential to making these models safer, fairer, and more effective.
Recital 12b	N/A – comment only	The text of the Act does not clarify the application of these

Neither the collaborative development of free and open-source AI components nor making them available on open repositories should constitute a placing on the market or putting into service. A commercial activity, within the understanding of making available on the market, might however be characterised by charging a price, with the exception of transactions between micro enterprises, for a free and open-source AI component but also by charging a price for technical support services, by providing a software platform through which the provider monetises other services, or by the use of personal data for reasons other than exclusively for improving the security, compatibility or interoperability of the software.		terms to foundation models. See proposed amendments to Article 3 above.
Recital 12c *The developers of free and open-source AI components should not be mandated under this Regulation to comply with requirements targeting the AI value chain and, in particular, not towards the provider that has used that free and open-source AI component. Developers of free and open-source AI components should however be encouraged to implement widely adopted documentation practices, such as model and data*	N/A – comment only	The text of the Act does not clarify how these principles apply to foundation model development. See proposed amendments to Article 28 above.

cards, as a way to accelerate information sharing along the AI value chain, allowing the promotion of trustworthy AI systems in the Union. *Recital 60* *Within the AI value chain multiple entities often supply tools and services but also components or processes that are then incorporated by the provider into the AI system, including in relation to data collection and pre-processing, model training, model retraining, model testing and evaluation, integration into software, or other aspects of model development. The involved entities may make their offering commercially available directly or indirectly, through interfaces, such as Application Programming Interfaces (API), and distributed under free and open source licenses but also more and more by AI workforce platforms, trained parameters resale, DIY kits to build models or the offering of paying access to a model serving architecture to develop and train models. In the light of this complexity of the AI value chain, all relevant third parties, in particular those that are involved in the development, sale and the commercial supply of software tools, components, pre-trained models or data incorporated into the AI system, or providers of network services,*		

should without compromising their own intellectual property rights or trade secrets, make available the required information, training or expertise and cooperate, as appropriate, with providers to enable their control over all compliance relevant aspects of the AI system that falls under this Regulation. To allow a cost-effective AI value chain governance, the level of control shall be explicitly disclosed by each third party that supplies the provider with a tool, service, component or process that is later incorporated by the provider into the AI system.		
Recital 60g *In light of the nature and complexity of the value chain for AI system, it is essential to clarify the role of actors contributing to the development of AI systems. There is significant uncertainty as to the way foundation models will evolve, both in terms of typology of models and in terms of self-governance. Therefore, it is essential to clarify the legal situation of providers of foundation models. Combined with their complexity and unexpected impact, the downstream AI provider's lack of control over the foundation model's development and the consequent power imbalance and in order to ensure a fair sharing of responsibilities*	N/A – comment only	The text of the Act does not clarify how these principles apply to foundation models. The body of the Act does not account for small, fine-tuned, or lower risk models. See proposed amendments to Article 28b above.

along the AI value chain, such models should be subject to proportionate and more specific requirements and obligations under this Regulation, namely foundation models should assess and mitigate possible risks and harms through appropriate design, testing and analysis, should implement data governance measures, including assessment of biases, and should comply with technical design requirements to ensure appropriate levels of performance, predictability, interpretability, corrigibility, safety and cybersecurity and should comply with environmental standards. These obligations should be accompanied by standards. Also, foundation models should have information obligations and prepare all necessary technical documentation for potential downstream providers to be able to comply with their obligations under this Regulation. Generative foundation models should ensure transparency about the fact the content is generated by an AI system, not by humans. These specific requirements and obligations do not amount to considering foundation models as high risk AI systems, but should guarantee that the objectives of this Regulation to ensure a high level of protection of fundamental rights, health		

and safety, environment, democracy and rule of law are achieved. Pre-trained models developed for a narrower, less general, more limited set of applications that cannot be adapted for a wide range of tasks such as simple multi-purpose AI systems should not be considered foundation models for the purposes of this Regulation, because of their greater interpretability which makes their behaviour less unpredictable.		