Supporting Open Source and Open Science in the EU AI Act

Executive Summary:

Open source, non-profit, and academic research and development play an essential role in the artificial intelligence (AI) ecosystem. Continuing to support and foster this open ecosystem will be paramount to ensuring that the technology serves all EU citizens on two main accounts:

- First, the values of sound research, reproducibility, and transparency fostered by open science are instrumental to the development of safe and accountable AI systems.
- Second, open source development can enable competition and innovation by new entrants and smaller players, including in the EU.

The AI Act holds promise to set a global precedent in regulating AI to address its risks while encouraging innovation. By supporting the blossoming open ecosystem approach to AI, the regulation has an important opportunity to further this goal through increased transparency and collaboration between stakeholders. Unfortunately, current proposals threaten to create impractical barriers to and disadvantages for contributors to this open ecosystem.

The undersigned organizations represent both commercial and nonprofit stakeholders in the open source AI ecosystem. Below, we make 5 concrete suggestions for how to ensure the AI Act works for open source:

1. Define AI components clearly
2. Clarify that collaborative development of open source AI components and making them available in public repositories does not subject developers to the requirements in the AI Act, building on and improving the Parliament text's Recitals 12a-c and Article 2(5e).
3. Support the AI Office's coordination and inclusive governance with the open source ecosystem, building on the Parliament's text.
4. Ensure the R&D exception is practical and effective, by permitting limited testing in real-world conditions, combining aspects of the Council's approach and an amended version of the Parliament's Article 2(5d)
5. Set proportional requirements for "foundation models," recognizing and distinctly treating different uses and development modalities, including open source approaches, tailoring the Parliament's Article 28b.

More detail on these suggestions is provided in section 4 at the end of this document, including specific text suggestions in the Annex. In sections 1 to 3, we explain the broader

open ecosystem of AI development; highlight how other EU policy instruments have recognized and adapted approaches suited to open source; and assess how the existing proposals for an AI Act would impact the open source ecosystem.

Signatories:

Creative Commons

EleutherAI

GitHub

Hugging Face

LAION

Open Future

0. Introduction

Open source, non-profit, and academic research and development play an essential role in the artificial intelligence (AI) ecosystem. This position paper describes current open ecosystems and practices in AI, analyzes the co-legislators' AI Act positions, and concludes with a series of recommendations to ensure that the final AI Act can support the open ecosystem to build safe, reliable, and beneficial AI technology.

Open source, open science, and open culture allow anyone to learn from and build on ideas and creations of the past, and to participate in the collaborative development and study of the latest technological advances by bringing in diverse perspectives and expertise. They've enabled knowledge commons like Wikipedia and yielded open source software—code that can be used, studied, distributed, and modified freely—that generates between €65-95 billion for EU GDP annually.

In the specific context of AI and Machine Learning (ML), recent advances of the technology are a direct consequence of an extensive ecosystem of open scientific research and open source development. The software used to train, deploy, and use new models all rely heavily on open source software, notably through the vibrant developer communities that support frameworks like PyTorch and TensorFlow. The major scientific innovations in model architecture and training paradigms of the last decade or so[1] have been made possible by the open exchange of ideas between academic and industrial researchers and developers of diverse backgrounds. This is also true for the majority of the evaluation benchmarks that shape the development of the field[2] and the research that furthers our understanding of the internal workings of those systems.[3]

Continuing to support and foster this open ecosystem going forward will be paramount to ensuring that the technology serves all EU citizens on two main accounts.

First, the values of sound research, reproducibility, and transparency fostered by open science are instrumental to the development of safe and accountable AI systems. Open and accessible sharing of the software, datasets, and models that make up AI systems allows for more widespread scrutiny and understanding of their capabilities and shortcomings—among academics, developers, regulators, and the public at large. This transparency, enabled by open licensing and direct access to AI components, supports research on the bias, safety, environmental, and security concerns posed by AI. It fosters more robust and inclusive mechanisms for AI accountability that can address the root

[1] E.g., the first general-purpose machine vision systems like AlexNet; the attention mechanism; language pre-training in natural language processing.
[2] E.g., GLUE, BigBench, Harness, openCLIP Benchmark
[3] E.g., InterpretML, AI Fairness 360, Fairlearn

causes of negative impacts of this technology, such as detecting and evaluating racial bias in facial recognition.

Second, open source development can enable competition and innovation by new entrants and smaller players, including in the EU. Open source development has disrupted narratives claiming that commercially successful AI systems could only be developed by very large companies in the US and China. High costs associated with training new AI models has traditionally been a significant barrier to competition in the AI market. Open collaborations, such as EleutherAI and BigScience, among many others, bring together hundreds of researchers across institutions to develop and share the resources and skills required to train these models. These efforts have shown that open collaboration can help overcome barriers to develop large models that are both competitive with those of private entities and more aligned with the principles that drive the EU AI Act. Once trained, these models can be fine-tuned by downstream providers at a fraction of the initial training cost. This new paradigm has shifted the market and promises to increase competition, offering developers a greater variety of AI tools that better suit their specific needs and reflect their values. In short, open source offers the EU another path, one aligned with European values.

The AI Act holds promise to set a global precedent in regulating AI to address its risks while encouraging innovation. **By supporting the blossoming open ecosystem approach to AI, the regulation has an important opportunity to further this goal through increased transparency and collaboration among diverse stakeholders.** That is not to say that openness and transparency are sufficient alone to ensure positive outcomes as the technology sees broader adoption. Rather, AI requires regulation that can mitigate risks by providing sufficient standards and oversight, verifying that systems are fit for purpose when they are put into service or placed on the market, and establishing clear liability and recourse for harms. In this paper, we make specific recommendations to balance needed risk mitigations with support for the open AI ecosystem.

1. The open ecosystem of AI development

1.1 Overview of collaborative ML development practices

1.1.1 Machine learning concepts and components

The AI systems that we interact with are an amalgamation of several components. At the core is a **model**, which takes an input to produce an output. In ML, the dominant AI paradigm today, developers select an architecture for the model's parameters and then set the weights associated with each parameter by training the model on **data**. When given a new input, the model then produces an output that reflects learned patterns from its training data; this is called "inference." Deployment **software**, which includes code required

to run inference and data pipelines, manages these inputs and outputs and what's done around them. Together the model and software code constitute an **AI system**.

Importantly, to understand how a given AI system works in practice, it is necessary to understand how it was trained. This stands in contrast to traditional software systems, where access to the source code is typically sufficient to explain and reproduce behaviors. The **training dataset**, **training algorithm**, **code used to train** the model, and **evaluation datasets** used to validate the development choices and quantify the model performance all impact how the system operates and are not immediately apparent from inspecting the final AI system. This raises the importance of **documentation** and information sharing, especially when knowledge of or access to those components may be required to evaluate a new aspect of an AI system, or to evaluate an AI system in a new deployment setting.

These components have become the building blocks of the AI systems we see today. At the moment of writing, leading software collaboration platform GitHub hosts nearly 1.8 million repositories focused on ML and 11+ million repositories of open data, and their contributors are diverse: among the 20 most popular generative AI projects on GitHub, nine belong to individuals, nine to academic organizations and startups, and two to large technology companies. Hugging Face, a leading platform for sharing open AI components, hosts over 250,000 models, 50,000 datasets, and 50,000 demos with their deployment software. Both platforms also feature and foster extensive documentation of the hosted models and datasets, following the recommendations of established scholarship on datasheets, data statements, and model cards[4]. Among those components, **pre-trained models** play a particularly important role in alleviating the significant costs, environmental impact and data requirements associated with training a new model for every new AI application. Of the open, pre-trained models,[5] several hundred fit the current AI Act definition of foundation models, all with **different languages, training distributions, modalities**, and fitness for adaptation to **different ranges of applications**. Many popular recent AI systems, and in particular those known as generative AI, rely on a two-step process, where a model is first pre-trained on a general-purpose objective to encode statistics about the world, and then adapted for a specific task. While creating the original pre-trained model still takes significant resources, subsequent modification can **drastically reduce compute required to run inference**, including via methods of distillation and quantization. Pre-trained models can then be **fine-tuned** for a specific task among a wide variety of applications, again at much lower cost, e.g., via Low-rank Adaptation (LoRA, QLoRA). The result is **modular**

[4] Datasheets document the motivations behind gathering a given dataset, as well as any data processing and impacted stakeholders. Model cards include disclosures about the model architecture and development, its intended and out-of-scope uses, evaluation metrics, training data, and ethical considerations.

[5] E.g., GPT-NeoX 20b, openCLIP ViT G/14

development, where pre-trained models benefit from multiple, stack-able improvements downstream.

By freely sharing all of these components, stakeholders with diverse skills and interests can collaborate not just on improving them, but also on building a better understanding of how they fit together and of their limitations. When these components are shared as open source software or with similar permissive licensing terms, they **also allow developers to build products** that benefit from all of this existing work and expertise by selecting the models that are best suited to their needs and values—with full control of and visibility into the development process, rather than trying to fit a single model into their product.

1.1.2. Examples of existing open, collaborative development

In order to better illustrate the importance of open sharing, we now review several efforts that showcase the modularity and value of collaborative development of ML systems. In particular, we outline how the different structure of open research encourages technological innovation that addresses a broad range of technical and social challenges.

For example, while most commercial models have a heavy (if not exclusive) focus on English, European research institutions and independent developers have done extensive work releasing models in their own languages, including but not limited to Spanish,[6] Basque,[7] French,[8] and Nordic languages.[9]

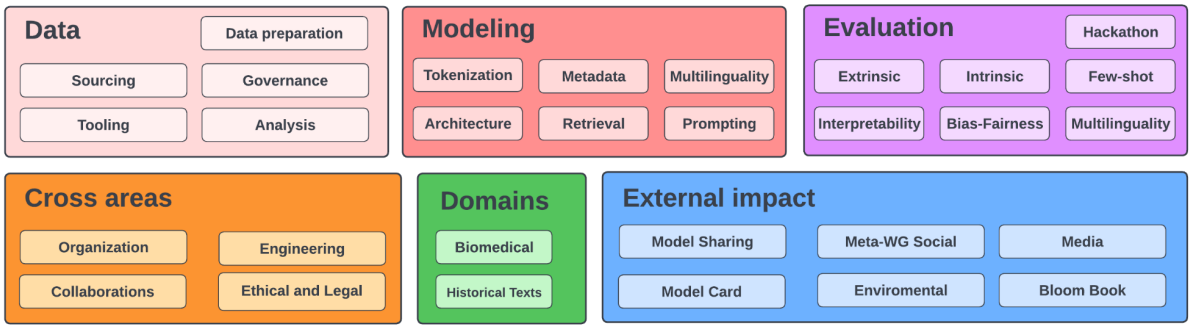
Open research also allows researchers with shared interests to pool resources across wide collaborations. While large AI models may be perceived as only the province of large companies, pooling resources allows others to develop similarly sized models that put a heavier emphasis on reusability, transparency, inclusion, or diversity. We review two such collaborations here: EleutherAI and BigScience.

EleutherAI is a community of AI researchers that started as a grassroots effort to create a publicly available state-of-the-art large language model (LLM). In two and a half years, it grew into a leading open source ML lab. In pursuit of open science, EleutherAI has created, documented, and publicly released every component necessary to pre-train a large language model. This began with The Pile, a thoroughly-documented 800GB model training dataset. **The construction of this corpus has since served as a guide for other open source projects**, such as the recently released Falcon model and its dataset. The organization has also released and maintains two open source code libraries that are crucial for contemporary AI development: one library for training LLMs from scratch, and one

[6] Barcelona Supercomputing Center
[7] University of the Basque Country
[8] INRIA
[9] Research Institutes of Sweden

library for [running model benchmarks in a standardized and reproducible fashion]. **These software building blocks are used by many researchers and developers**, not just EleutherAI. They have enabled [foundational research] that would have been impossible without EleutherAI's fully open releases and numerous commercial products that integrate their components.

Owing to compute research grants, EleutherAI has been able to release over two dozen pre-trained LLMs. The most recent EleutherAI release is a set of 16 LLMs, with 154 intermediate checkpoints per model (which are models in themselves, but at different stages of training). This project — called the Pythia suite — is **designed for scientific study of language model behavior**, as each model was trained on the same data in the same order, and all models have the same architecture. This setup makes it possible to answer precise questions about, among other things, models' ability to [reproduce material] from their training dataset, [model and dataset auditing], and [relationship between bias in the training data and model outputs]. These highlighted research topics have implications for AI policy in areas such as fairness, copyright, and privacy. EleutherAI's open release of models and AI components allows other actors to independently adapt them to their research and commercial needs, including fine-tuning pre-trained LLMs into multiple European languages (such as [Spanish] and [German]), or training models from scratch (for example, [Romanian] and [Slovak] LLMs).



Working groups in the **BigScience** project illustrating the many different areas of study and development for a single Large Language Model relying on direct access to its development

[The BigScience workshop] was a year-long project organized around the training of an open and **responsible multilingual LLM** - with several [European languages covered in the language set], including regional languages. The BigScience project was enabled by a resource grant on the French [public computing infrastructure Jean Zay], supported by companies including Hugging Face, and gathered over 1200 multidisciplinary researchers from around the world to work on all of the many components that make up an AI system. The [organization of the project into over 30 working groups], each working on different research questions connected to the design and understanding of the final model while benefiting from direct access to the other components, illustrates **the need to facilitate**

direct access to AI components across institutions and across disciplines. The open nature of this collaboration allowed participation from a wide range of expertise relevant to the **regulatory context and social impact of the technology**. The project was able to both use existing open tools when available and to release a wide range of new components to support further research; including new domain-focused datasets, new data governance methodologies and tools, data processing software, etc. In particular, the release of the BLOOM model under an open Responsible AI License has also allowed external researchers to adapt the model to new languages and to integrate it into a chat-based AI system supporting most of the model's training languages.

1.2 Supporting EU values through open access to AI components and systems

The previous section provided an overview of ways in which the open research and development of AI systems differs from that of proprietary systems, outlining its contribution to making ML more transparent, inclusive, modular, and diverse. Next, we argue that **broad access to functional AI systems and open general-purpose models is both a necessary and a valuable part of open and accountable development, especially when their components are similarly openly available.**

For example, LAION is a non-profit research organization that works to openly release large AI datasets, components, and models. They're perhaps most well-known for working on LAION-5B: describing and open-sourcing a pipeline for composing large-scale image-text datasets to enable reproducible studies on language-vision learning. Using such reproducible datasets and open, standardized, reproducible benchmarks, LAION has trained, studied and released a series of openCLIP models — foundation language-vision models that show strong capability to transfer across various important tasks. LAION has subsequently received the Outstanding Paper Award at NeurIPS 2022 for their work on language-vision models (LVMs) training datasets.

Open access to these systems is particularly relevant in the context of the analysis of social biases in image generation AI systems. The Lensa application uses AI to transform personal photos into glamorized illustrations of the pictured individuals. Reporting has identified biases in the application, namely that it produced sexualized images of women. The application uses an openly available text-to-image model, Stable Diffusion, which was trained on the LAION 5B dataset. Thanks to open access to the training dataset, along with open source tools that include a search engine, making it easier to explore for users with limited resources, the reporter was able to identify specific images in the training dataset that promote this behavior. Since then, further research using the open access to this dataset has explored the particular role of scale in triggering these harmful behaviors. While proprietary systems are known to showcase similar biases as their open

counterparts, the lack of broad access to their components inherently limits most stakeholder's ability to gain meaningful understanding of the mechanisms that trigger these biases or of the efficacy of steps taken to mitigate them. It is precisely LAION's open release of training data that allows for this public scrutiny.

Open research projects have actively taken steps to ensure transparency and support evaluation. For instance, the BLOOM model was released with a search engine on its training dataset to allow users to analyze issues such as data contamination, factuality, or privacy risks. Direct access to open source LLMs and LVMs along with their components has allowed researchers to clear up misunderstandings about initially reported performance results in a way that would have been impossible without open access to the exact models, deployment, and evaluation software. In contrast, erroneous claims about a model's performance[10] can be significantly harder to disprove when access to the model and its components is restricted—putting into question the validity of self-reported results and descriptions of model capabilities without sufficient external validation.

Providing access to well-documented and transparent demos of systems can also be critical to supporting beneficial open systems. Supporting limited real-world interactions significantly increases the diversity of stakeholders who can participate in those evaluations. The availability of suitable demos for systems also facilitates broader red-teaming efforts, automatic evaluations, and crowd-sourced human evaluation, and scrutiny from stakeholders with less technical resources. All of this contributes to more systematic evaluation of AI systems, namely their strengths, risks, and areas for improvement. Broadly speaking, interactive demos accompanying models are increasingly becoming an integral part of good documentation for researchers as well as developers. While model cards can provide fixed, general information, interactive demos enable users to understand how a model will behave in a new context, allowing them to more easily identify specific limitations or assess a model's fitness for a specific application. In turn, it can help people make more informed decisions about whether and how to use specific AI components and systems.

Finally, open access to pre-trained models has supported work to make training and inference with LLMs and LVMs more resource-efficient, including driving the adoption of parameter-efficient fine-tuning methods like LoRA. Developers can cheaply patch models to adapt to their own use cases, taking advantage of open source libraries to run large models efficiently on CPUs. Between the lower cost of building on pre-trained models, the ability to better track the cost of training, and the demand for more compute-efficient technology from typically less-resourced participants, **open development and sharing of AI systems can play a major role in fostering more energy- and carbon-efficient AI innovation.**

[10] E.g., about a model's ability to achieve "zero-shot translations" or high performance on academic tests, or over-fitting on specific versions of tests.

Benefitting from a public good requires reasonably free access to that good in the first place. It is therefore imperative for the EU that AI Act requirements not disproportionately burden open source and science AI development. We elaborate on how this can be achieved in the following sections.

2. Learning from policy on open source software

Although the open ML ecosystem is only now beginning to develop, over three decades of experience with the open source software ecosystem can provide valuable insights. Today, open source software is ubiquitous and extensive: it is in nearly all software and, for a given software stack, generally makes up over three-quarters of the constituent components. In short, this global public good has become essential to our digital infrastructure.

The open source software ecosystem has developed thanks to contributions of countless developers from around the world. It has also been enabled by software licenses that disclaim warranty and liability for contributions. Individual contributors may share an improvement or security fix with upstream project maintainers, and those maintainers can choose to update their project and in turn promulgate that update downstream to all users of the software. In this way, individual improvements can be evaluated and disseminated to make more effective software.

Public policy has protected community collaboration on open source, including in the EU. The 2019 Copyright Directive recognized the importance of protecting open source software development platforms from expectations to proactively filter uploads, and so expressly exempted these platforms. When proposed by the Commission in 2022, both the Cyber Resilience Act and the Product Liability Directive included exemptions for open source software, and both Parliament and Council have offered draft positions that build on these exemptions.

Ongoing legislative work on the Cyber Resilience Act and the Productive Liability Directive is worth detailing. Although not yet fixed, several themes can be observed in recent texts:

- First, the proposals aim to **exempt open source software that is provided outside the context of a commercial transaction.** This would be the case for both the source code objects and the running software, so a non-profit would be free to provide, for example, a productivity application for end-use without complying with expectations for monetized products.
- Second, **proposals are agnostic as to the context of development when determining whether the software should be subject to the regulation.** That is, corporate developers are free to collaborate on the development of open source software without being subject to the product-oriented regulations. Similarly, the

services that support open source collaboration, including public repositories, are not subject to requirements for distributors of products.

- Third, **the open source community has no obligation to collaborate downstream**: if a manufacturer wishes to integrate an open source software component into a product, it is the manufacturer that faces obligations to ensure that the final product is compliant. These three themes – non-commercial supply is exempted, development is exempted, and obligations fall downstream – reflect norms that have enabled open source collaboration to flourish.

Decades of open source experience should inform the AI Act as should these parallel legislative files. However, it is worth noting that definitions of open source AI are not yet fixed and will have to grapple with the complex interactions between the different components of an AI system. As AI model development has moved from expensive training from scratch to further training of open pre-trained models, the openness of the code, documentation, model, and meaningful transparency about the training data[11] empower anyone to use, study, modify, and distribute the AI system as they would open source software.

3. Assessing the proposals

3.1 Commission proposal

As originally proposed by the Commission, the AI Act governed only high risk AI systems placed on the market or put into service. This scope provided clarity that (open source) AI components, including models, were not subject to the regulation unless they are integrated in high risk AI systems as defined by the Act.

3.2 Parliament report

The Parliament position provides helpful clarity on open source AI components in Recitals 12a-c. Recital 12a clarifies open source AI components that are not supplied in a commercial transaction, excluding micro-enterprise transactions, are exempted from the Act's provisions. Non-commercial deployment of banned, high-risk, or AI systems facing transparency obligations are within scope, reducing risk of loopholes. Recital 12b clarifies that collaborative development of AI components or hosting them on an open repository does not constitute making them available on the market nor putting them into service. Recital 12c clarifies that open source developers face no obligation to collaborate downstream if a provider wishes to integrate their open source component into an application. These are all helpful provisions for the open source ecosystem.

[11] To be meaningful, transparency should include the provenance, composition, and processing steps associated with the training data.

However, Article 2(5e) includes a modified version of Recital 12a which introduces ambiguity and challenges. Specifically, the final sentence “This exemption shall not apply to foundation models as defined in Art 3,” and thus appears to exclude foundation models from the open source exemptions discussed above. Similarly concerning, Article 2(5d) appears to impede the beneficial practice of offering demos of AI systems with the text “The testing in real world conditions shall not be covered by this exemption.”

Requirements on AI models should be calibrated to their risks, crafted in a way that helps address those risks, and be capable of being implemented. As scoped, the newly introduced Article 28b on foundation models presents a number of challenges for a foundation model developer without significant financial resources and institutional backing, such as volunteer efforts that release open source models. This is despite the many benefits of open AI research efforts in relation to transparency and documentation.

Requirements for all foundation models should focus on the mitigation of identified risks, transparency, and adequate documentation. Detailed documentation is in fact necessary for any successful open source project, since it makes it easy to welcome new contributors and instruct potential users.

Beyond this, many proposed provisions in Article 28b are not practical for open source foundation model providers and would impede beneficial development. Key issues include:

- The biggest obstacle would be establishing a quality management system as required in article 28b(2(f)), as that typically requires (1) a final product that is being aimed at and therefore dictates the standards acceptable for a project, and (2) dedicated staff with knowledge of both the law and relevant technological practices. This is something that a typical volunteer project does not have the financial and human resources for.
- The involvement of independent auditors (28b(2(a))) could be interpreted to require contracting third-party specialists, which is costly and not necessary to mitigate the risks associated with many foundation models. Moreover, openly available foundation models and AI components already allow for great scrutiny from independent experts. Indeed, they are frequently more accessible to auditors than closed systems, and this accessibility is built into the philosophy underpinning open science and open source software.
- Collaborative open development efforts usually do not control the computing clusters on which models are trained and hosted. This has implications for Article 28b(2(d)) — accurate measuring and logging of resource consumption can only be performed by the compute provider.
- It is unclear who could be assigned the responsibility to keep documentation for 10 years after a foundation model (as required by 28b(3)) in the context of decentralized development that lacks institutional backing. Open source collaborations are often

ad hoc without a formal organization. For nearly three years prior to becoming a legal entity, EleutherAI was simply a volunteer group of researchers.

Ultimately, the difficulties for open source foundation model providers stem from how open development differs in methods and goals from the large-scale, commercial and closed models that have become most familiar to people and appear to be the focus of the Parliament's text. Open foundation models like those developed by EleutherAI, LAION or BigScience's BLOOM are first and foremost research artifacts and are intended for downstream use only after specific adaptation. In contrast, the majority of commercially available language model APIs rely on a private model that has already been fine-tuned to follow user instructions and simulate conversation. They also involve several additional technical layers and significant computational resources to support their use by potentially millions of users and integration in customer-facing applications. Ultimately, API offerings are quite different from models developed and shared by the open source community.

Due to the differences in resources and release strategies, we advocate below for proportionality — the requirements for foundation models should fit the scope and impact of a given AI system and the developer's position in the industry and the AI value chain.

3.3 Council position

The Council's common position introduced the concept of "general purpose AI systems" that included text, "irrespective of whether the general purpose AI system is placed on the market or put into service as a pre-trained model and whether further fine-tuning of the model is to be performed by the user of the general purpose AI system" (Article 4a(2)). These AI components could be in scope, "including as open source software" (Article 3(1b)). By applying many of the requirements for high-risk services to open source projects, these provisions create similar challenges to those discussed in the context of the Parliament's text on "foundation models." It appears to have been the intent of the Council with its proposal to address the AI value-chain with systems that could be adapted for downstream use. In order to avoid confusion, we recommend focusing on an improved definition of the "foundation model" concept in lieu of general purpose systems in the final negotiation.

In contrast, the Council's Article 2(6-7) provides useful clarity that AI systems designed and put into service for scientific research and development are exempt from the Regulation as is all research and development activity regarding AI systems. These provisions offered helpful clarity that the open, collaborative development of AI components would not be considered placing on the market or putting into service and thus not trigger obligations under the Act. These provisions would also enable the limited demo deployment of AI systems for research purposes.

4. Recommendations

In this final section, we provide five recommendations for the Trilogue. These aim to ensure that open AI development practices are not confronted with obligations that are structurally impractical to comply with or that would be otherwise counterproductive. Overbroad obligations drafted with closed and proprietary AI development in mind threaten to disadvantage the open AI ecosystem. These recommendations offer amendments to the European Parliament's position as its position is both the most recent and has the most detailed requirements on sub-components of AI systems.

I. Define AI components

The Parliament position refers to an "AI component" without providing a specific definition. As a result, it is unclear which specific elements should be regarded as "components" in the context of the draft and thus what is subject to corresponding obligations under the Act.

II. Clarify that collaborative development of open source AI components and making them available in public repositories does not subject developers to the Act

The Parliament position provides helpful clarity on open source AI components in Recitals 12a-c. Open source software research and development follows a community-driven approach, where individuals contribute their expertise and collectively improve the code. By making the AI components freely available, developers encourage collaboration, knowledge sharing, and innovation within the AI community. Collaborative development of open source AI components entails making them available for others to study and modify. This process includes sharing the components on open repositories. None of these activities should be misconstrued as commercial activities subject to market regulations and considered as placing the components on the market or putting them into service. Neither should it trigger requirements targeting the AI value chain in Article 28.

III. Support the AI Office's coordination and inclusive governance with the open source ecosystem

Regulating a fast changing technology like AI demands regulatory capacity. Ensuring that a general purpose technology with vast potential to impact society, like AI, is governed democratically also requires diverse participation. Thankfully on both fronts, Parliament has proposed bolstering the Commission and Council's AI Board into an AI Office. We believe the AI Office can play an important role in coordinating work across national regulators, engaging with international partners and ensuring effective implementation of the Act. The AI Office and its advisory forum can also play an important role in providing regulatory

clarity and inclusive deliberation on how the evolving open source AI ecosystem may be subject to provisions of the AI Act. We recommend that the final Act include Parliament text for an AI Office with some modest amendments.

IV. Ensure the R&D exception is practical and effective, by permitting limited testing in real-world conditions

Best practice in AI research and development today commonly requires access to operational AI systems for various purposes including testing and red-teaming. Although AI components can be made publicly accessible as static code, providing them as interactive systems for exploration and demonstrations has many additional benefits. However, the current version of the text from Parliament appears to impede any testing under real-world conditions. While we recognize that the purpose of this provision is to avoid creating a legal loophole, we believe that, if adopted in the current form, it will significantly impede research and development. Understanding the performance of AI systems, identifying potential weaknesses, and improving their functionality all depend on the ability to test them in real-world scenarios. The proposed clause may therefore prevent researchers and developers from learning important lessons and enhancing the robustness of AI systems.

V. Set proportional requirements for “foundation models,” recognizing and distinctly treating different uses and development modalities, including open source approaches

We appreciate the intent of the text to provide appropriate safeguards for technologies that can support a variety of downstream tasks, as reflected in the Parliament's proposed requirements for “foundation models” and the Council's earlier focus on “general purpose AI.” We also acknowledge that a number of the obligations that the Parliament’s text imposes on “foundation models” are both justified and structurally aligned with the principles of open source development.

However, the current one-size-fits-all approach requiring full control of the development chain creates insurmountable barriers for participants in the open AI ecosystem. In turn, these provisions would lead to further concentration of development, deployment and understanding of this technology, with concerning implications for European competitiveness. We have outlined the challenges this could pose to the development of operating systems for AI systems in Section 3.

Instead, the Act ought to recognize and distinctly treat different uses and development modalities. Creating these distinctions is consistent with the overall approach of the Act, which focuses on tailoring requirements to fit particular risks, accounting for fundamental rights and other concerns, and— in particular the goals of Title V—measures to support

innovation, including “measures to reduce the regulatory burden on SMEs and start-ups” and support “small-scale providers.”

We believe that a more tailored framework would be able to address these concerns without undermining the regulatory objectives of Article 28b. This can be achieved by **distinguishing between a set of baseline obligations that should apply to all foundation models put into service, while limiting some more far-reaching obligations to a subset of models that are commercially deployed or reach a particular threshold meriting additional scrutiny.** In addition, **the AI Office should be authorized to periodically review both the nature and the threshold for application of these additional obligations.**

Baseline requirements should apply to all foundation models that are put into service or made available on the market, and should ensure meaningful transparency, data governance, technical documentation, and risk assessment. Mapped on the text proposed by Parliament, this could concern the following requirements of Article 28b: (2a), (2b), (2e), and (4).

The remaining subset of requirements contained in Article 28b of the Parliament's text would only apply to those models made available on the market, or that are put into service and exceed a certain threshold, based on further analysis of foundation models. The AI Office should set the threshold based on current research and the state of the art. In addition, the AI Office should be empowered to periodically review the nature of the requirements imposed on foundation models.

This tiered approach would address the risk that the current one-size-fits-all approach creates impossible barriers for entities developing models that are not deployed at scale. It would avoid a situation where open source developers of foundation models face structural barriers, do not have the resources to handle the regulatory burden, and do not have full control of their entire development chain. Thus tailored, the AI Act can encourage beneficial AI development and competition in the single market and beyond.

ANNEX - Compilation of Specific Proposed Changes to Text

I. Define AI components clearly

Article 3

European Parliament	Proposed Language
	1f - 'AI component' means any software element required to build and operate an AI system, including software code, training data, a model and its weights.
Justification: Recitals 12a-c and Art 2(5e) reference "AI components" while leaving the term undefined. We recommend defining the term so as to incorporate any component required for a functioning AI system and to do so expressly.	

II. Clarify that collaborative development of open source AI components and making them available in public repositories does not subject developers to the requirements in the AI Act, building on and improving the Parliament text's Recitals 12a-c and Article 2(5e).

Article 2

European Parliament	Proposed Language
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 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.	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 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 ***to the extent they are placed on the market or put into service.***
Justification: Recitals 12b states plainly that "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," and Recital 12c similarly notes that Article 28 should not apply to creators of OSS components. The "foundation model" requirements in Article 28b only apply to models that have been "put into service" or "placed on the market," and Recital 12b would seem to exclude OSS from these requirements as well. However, this is left ambiguous in the original Article 5 text.	

III. Support the AI Office's coordination and inclusive governance with the open source ecosystem, building on the Parliament's text.

Article 56b

European Parliament	Proposed Language
	t - provide monitoring of and interpretative guidance for the ever evolving open source ecosystem and applicability of provisions of the AI Act.
Justification: The list of tasks for the AI Office identified in Article 56b includes a range of helpful activities, including interpretative guidance and monitoring for the providers of foundation models and evolving AI value chains. We recommend adding an additional task expressly focused on the open source ecosystem to ensure regulatory clarity from the AI Act and to monitor its impact on open source development.	

Article 58

European Parliament	Proposed Language
2 - The membership of the advisory forum shall represent a balanced selection of stakeholders, including industry, start-ups, SMEs, civil society, the social partners and academia. The membership of the advisory forum shall be balanced with regard to commercial and non-commercial interests and, within the category of commercial interests, with regards to SMEs and other undertakings.	2 - The membership of the advisory forum shall represent a balanced selection of stakeholders, including industry, start-ups, SMEs, civil society, ***open source developers,*** the social partners and academia. The membership of the advisory forum shall be balanced with regard to commercial and non-commercial interests and, within the category of commercial interests, with regards to SMEs and other undertakings.
Justification: The advisory forum of the AI Office provides important multistakeholder input into the enforcement of the AI Act. We recommend expressly including open source researchers as a stakeholder group so as to support responsible open source development within the EU.	

Article 82b

European Parliament	Proposed Language
1h - The practical implementation of Article 12, Article 28b on environmental impact of foundation models and Annex IV 3(b), particularly the measurement and logging methods to enable calculations and reporting of the environmental impact of systems to	1h - The practical implementation of Article 12, Article 28b on environmental impact of foundation models and Annex IV 3(b), particularly the measurement and logging methods to enable calculations and reporting of the environmental impact of systems to

comply with the obligations in this Regulation, including carbon footprint and energy efficiency, taking into account state-of-the-art methods and economies of scale. When issuing such guidelines, the Commission shall pay particular attention to the needs of SMEs including start-ups, local public authorities and sectors most likely to be affected by this Regulation.	comply with the obligations in this Regulation, including carbon footprint and energy efficiency, taking into account state-of-the-art methods and economies of scale. When issuing such guidelines, the Commission shall pay particular attention to the needs of SMEs including start-ups, ***open source developers,*** local public authorities and sectors most likely to be affected by this Regulation.
Justification: Guidelines from the Commission on the implementation of the AI Act, in consultation with the AI Office, as described in Article 82b, will provide needed clarity to aid compliance. Open source researchers should be expressly included as an audience for these guidelines in order for the Act to support responsible open source AI innovation.	

IV. Ensure the R&D exception is practical and effective, by permitting limited testing in real-world conditions, combining aspects of the Council's approach and an amended version of the Parliament's Article 2(5d)

Article 2

European Parliament	Proposed Language
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 may adopt delegated acts in accordance with Article 73 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 at coordinating its application by the national supervisory	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, ***except where such testing is done on a limited scale, with sufficient documentation and transparency to users***. The Commission is empowered to ~~***may***~~ adopt delegated acts in accordance with Article 73 to specify this exemption to prevent its existing and potential abuse. The AI Office shall provide guidance on the governance of research and development

authorities.	pursuant to Article 56, also aiming at coordinating its application by the national supervisory authorities.
Justification: Research and development (R&D) is crucial to the development of beneficial, trustworthy AI systems. The Act should recognize that some real world testing, including preliminary exploration of a model's appropriateness to specific deployment conditions and allowing scrutiny and evaluation by relevant civil society organisations outside of the development chain, can be necessary and appropriate for R&D. In particular, non-profit and research organizations may place a foundation model into service for this limited research purpose to beneficial ends. We recommend expressly permitting testing in real-world conditions at limited scale. The precise bounds of this permission can be guided by the AI Office pursuant to Art 56.	

V. Set proportional requirements for "foundation models," recognizing and distinctly treating different uses and development modalities, including open source approaches, tailoring the Parliament's Article 28b.

Article 28b

European Parliament	Proposed Language
2 - For the purpose of paragraph 1, the provider of a foundation model shall: (a) demonstrate through appropriate design, testing and analysis that 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;	2 - For the purpose of paragraph 1, the provider of foundation model shall: (a) demonstrate through appropriate design, testing and analysis that 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. This shall be without prejudice to relevant existing Union and national law and this obligation shall not apply before the standards referred to in Article 40 are published. They 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 paragraph C. When fulfilling those requirements, the	***(c) ~~(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.; ***3 - Additional measures shall apply to the provider of a foundation model that is put on the market, or that exceeds a threshold to be defined by the AI Office:*** ***(a) ~~2(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; ***(b) ~~2(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. This shall be without prejudice to relevant existing Union and national law and this obligation shall not apply before the standards referred to in Article 40 are published. They shall be designed with capabilities enabling the measurement and logging of the consumption of energy and resources **based on available information provided by compute provider**, and, where technically feasible, other environmental impact the deployment and use of the systems may have over their entire lifecycle; ***(c) ~~2(f)~~*** establish a quality management system to ensure and document compliance with this Article, with the possibility to experiment in fulfilling this requirement. ***(d) ~~2(g)~~*** register that foundation model in 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 notably 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 1(c) at the disposal of the national competent authorities;	EU database referred to in Article 60, in accordance with the instructions outlined in Annex VIII paragraph C. ***(e) ~~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 above at the disposal of the national competent authorities; **...** When fulfilling the above 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 notably in benchmarking guidance and capabilities referred to in Article 58a. ***The AI Office shall produce interpretative guidance to inform proportionate application of these provisions.***
Justification: We acknowledge that a number of the obligations the EP text imposes on foundation models are both justified and structurally very close to the principles underpinning the open source development model. However, the current one-size-fits-all approach risks creating insurmountable barriers for participants in the open source AI ecosystem. Instead, the Act ought to recognize and distinctly treat different uses and development modalities. This can be achieved by distinguishing between a set of baseline obligations that should apply to all foundation models, while limiting some more far-reaching obligations to a subset of models that are commercially deployed or meet certain thresholds worthy of further scrutiny, to be devised by the AI Office. In addition, the AI Office should be authorized to periodically review both the nature and the threshold for application of these additional obligations.	