

/ Recommendations by academics and activists

The AI Act and General Purpose AI: Charting a path forward

September 2023*

With the **EU's Artificial Intelligence Act (AIA)** entering its final phase of negotiations, this paper spotlights certain recommendations on how to govern systems that can be deployed for a variety of purposes ("General Purpose AI" - GPAI) and that generate content ("Generative AI" - GenAI).

The recommendations result from a workshop among academics and representatives of civil society organizations held in July 2023 and co-organized by *AlgorithmWatch* and the *AI, Media & Democracy Lab of the University of Amsterdam*.¹ Given their specific focus on GPAI, these recommendations are not comprehensive but should be regarded as complementary to other recommendations on the AIA.

* Update as of 15 September 2023, please refer to <https://algorithmwatch.org/en/ai-act-general-purpose-ai/>

¹ The workshop was held in July 2023 at the University of Amsterdam. A more comprehensive version of the recommendations will be published in an upcoming article. We are indebted to the **participants** for their valuable input during and after the workshop (participation in the workshop does not equal endorsement of all the recommendations outlined below): *Dr. Bettina Berendt (Professor for Internet and Society, TU Berlin)*, *Dr. Ian Brown (Consultant; Visiting Professor at the Centre for Technology and Society at Fundação Getulio Vargas Law School, Rio de Janeiro)*, *Dr. Nick Diakopoulos (Professor in Communication Studies and Computer Science (by courtesy) at Northwestern University)*, *Tim de Jonge (PhD candidate, Radboud University)*, *Christina Elmer (Professor for Digital Journalism/Datajournalism, University of Dortmund)*, *Dr. Natali Helberger (Distinguished University Professor Law & Digital Technology, University of Amsterdam)*, *Clara Helming (Senior Policy & Advocacy Manager, AlgorithmWatch)*, *Karolina Iwańska (Digital Civic Space Advisor, European Center for Not-for-Profit Law)*, *Dr. Frauke Kreuter (Professor for Statistics and Data Science, LMU Munich)*, *Dr. Laurens Naudts (Postdoctoral Researcher in Law, University of Amsterdam)*, *Liliane Obrecht (PhD Candidate in Law, University of Basel)*, *Dr. des. Angela Müller (Head of Policy & Advocacy, AlgorithmWatch)*, *Estelle Pannatier (Policy & Advocacy Manager, AlgorithmWatch CH)*, *Dr. Stanislaw Piasecki (Postdoctoral Researcher in Law, University of Amsterdam)*, *Dr. João Quintais (Assistant Professor in Information Law, University of Amsterdam)*, *Matthias Spielkamp (Founder & Executive Director, AlgorithmWatch)*, *Dr. Daniel Oberski (Professor in Health Data Science, Utrecht University)*, *Dr. Ot van Daalen (Lawyer; Lecturer and Researcher in Information Law, University of Amsterdam)*, *Kilian Vieth-Ditlmann (Deputy Team Lead Policy & Advocacy, AlgorithmWatch)*, *Dr. Sophie Weerts (Associate Professor in Public Law, University of Lausanne)*, *Dr. Frederik Zuiderveen Borgesius (Professor of ICT and Law, Radboud University)*. In addition, we thank *Nikolett Aszódi (Policy & Advocacy Manager, AlgorithmWatch)*, *Paul Keller (Director of Policy, Open Future)*, and *Alex Tarkowski (Director of Strategy, Open Future)* for their written feedback on the workshop outcomes.

1. Clarifying Definitions

Clear and distinct definitions are essential in providing **legal certainty**. Technology-neutral definitions (for key concepts such as “Foundation Model,” “Generative AI,” “General Purpose AI”) are preferred as they avoid loopholes based on technical specificities and are more future-proof.

That said, the AIA must:

- contain definitions that are **clear, distinguishable** and **avoid redundancies**. e.g. between “Foundation Model” and “General Purpose AI”, which currently seem to overlap. “Generative AI” should be defined as a sub-category of GPAI (for systems that automatically generate content such as text, video, or images);²
- rework definitions so that they can be **objectively assessed without reference to a provider’s intent** or knowledge of internal engineering processes. References to intention in Art. 3(1b) [CNL]³ and Art. 3 (1d) [EP]⁴ should be removed;
- clarify what constitutes a “**substantial modification**” of a GPAI system that could require a reassessment of a system’s risk level;
- **clarify definitions of actors** along the complex AI value chain (provider, deployer/user, affected persons) to allow clear attribution of respective obligations;
- ensure that technologies claiming regulatory exemption and/or special treatment as “**open source**” systems be released under licenses aligned with the Open Source Initiative’s (OSI) definition, and that such OSI licenses apply to a minimum set of components (models, weights, training data, etc.).

2. Addressing Complexity, Scale, and Power Asymmetries

Three key aspects mark GPAI systems. First, the **value creation chain and the network of actors behind them are typically highly complex** – and thereby more obscure to the people affected and to society at large. Second, **power asymmetries** in the downstream relationships further exacerbate this. Third, the **scale** of the model, of extractive practices and of adoption – and thereby the number of people potentially affected – is typically higher compared to other AI systems.

² Throughout this paper, the term “GPAI” is used for both of what is referred to as “foundation models” and “GPAI” in the official mandates on the AIA. “Generative AI” is defined as a subclass of GPAI.

³ [CNL] = Mandate of Council of the European Union on AIA, (6 December 2022)

https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:ST_15698_2022_INIT.

⁴ [EP] = Mandate of European Parliament on AIA, (14 June 2023)

https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html.

/ Managing Downstream Relationships

In the case of GPAI systems, providers often hold the power to impose their contractual terms onto downstream deployers, further expanding their influence over how fundamental rights and systemic societal risks will materialize and be addressed.

To ensure responsible governance of downstream relationships and consider power asymmetries, the AIA must:

- include mechanisms of regulatory scrutiny regarding **fairness, quality and adequacy of contractual terms and instructions** and ensure that these terms or instructions not be abused by providers to unilaterally exclude all responsibility (Art. 4c. [CNL], which foresees such unilateral exclusion of high-risk uses, provides a dangerous loophole and should be deleted);
- extend Art. 28 (2a) and Art. 28a AIA [EP] to contracts for foundation models / GPAI;
- contain an obligation of **mutual assistance and cooperation** in Art. 28(3) AIA [EP], along the lines of Art. 4b(5) [CNL] but not limited to high-risk areas;
- require **providers** to create an easy, accessible possibility for **incident reporting** to inform continuous impact/risk assessment;
- require **deployers** to share with providers and other downstream users, including individuals affected, **information** needed for compliance with the AIA;
- consider the need to **differentiate between larger and smaller deployers**;
- be informed by more expertise on **open source models** and their specific risks in order to create evidence-based requirements governing their development and deployment.

/ Fairness of Extractive Practices

The value creation chain behind GPAI, especially GenAI, which relies on huge amounts of training data and computing power, comes with enhanced risks for injustices caused by **discrimination, environmental impact, workers' maltreatment, as well as copyright and data protection infringements.**

The AIA should subject operators along the value chain to strict due diligence obligations to systematically assess and mitigate these risks.

Specifically, the AIA should

- require GPAI providers and deployers to conduct a **fundamental rights impact assessment** (see below);
- require providers and deployers of GPAI to conduct an **environmental impact assessment**, to make the results of such an assessment transparent, and to minimize, by means of best effort obligations, the environmental impact during all phases of a system's life cycle, similar to Art. 4a(1f), Art. 28b(2) and Art. 29a [EP]);

- include provisions foreseen in Art. 28b 2(a), 2(b) and 4 [EP], but adding an explicit reference to **workers rights**;
- ensure adequate and effective **data protection** in relation to GPAI systems;
- ensure that the **rights of creators** are respected by GPAI, especially GenAI, as foreseen in Art. Art. 28b(4b, 4c) [EP], but in addition specifying transparency obligations in order to make them operationalizable (i.e. machine-readable format, easy-to-use opt-out, complemented by access rights and a reversal of the burden of proof) and clarifying vague notions (such as "sufficiently detailed summary" or "generally acknowledged state of the art") for example in an implementing act.

3. Avoiding Accountability Gaps along the Value Chain

/ Division of Responsibilities

The complex value creation chain and network of actors behind GPAI also make responsibilities harder to assign – they cannot be allocated solely to either the provider or the deployer. This increases the **risk of accountability gaps**.

The AIA must ensure a meaningful division of responsibilities among operators through:

- a **general principle on accountability**, highlighting shared responsibilities throughout the value chain, which could be added in Art. 4a(1) and 4a(2) [EP];
- obligations of **mutual assistance**, including sharing of information and instructions necessary to comply with this regulation.

/ Rights of Affected Persons

Especially in light of the above-mentioned risk of accountability gaps in the context of GPAI, it is crucial that **natural persons affected have access to rights and are enabled to enforce them**.

To enable natural persons affected to exercise their rights, the AIA should include:

- in its Art. 52, obligations to provide general and contextual **information** to affected persons, in an easy and accessible, visible and timely manner, considering specificities and vulnerabilities of the affected person, as well as their mode of interaction, on:
 - the envisaged consequences and risks of their interaction with the system;

- the mitigation strategies operators have implemented to address those risks;
- available remedies (in case these remedies – including substantive rights – are provided by other legal instruments, information of their existence and applicability).
- **enhanced means for contestation**, including:
 - a right to an explanation about a decision taken by or with support of GPAI;
 - a right to consult with a human;
 - a right to an effective remedy where the persons consider that their rights have been infringed as a result of non-compliance with the AIA;⁵
 - a right for persons to flag and notify authorities in cases of potentially non-compliant AI systems or deployments;
 - a right to object against the application of such systems to them;
 - a right not to have their personal data used for further training as a result of their interaction with GPAI and systems built through them.
 - ensuring that natural persons affected by a GPAI system can exercise these rights against *any operator* in the value chain.
- means to **collectively contest and scrutinize GPAI systems** by
 - allowing civil society organizations to demand investigations and flag potentially non-compliant systems or deployment;
 - directly and explicitly foreseeing collective redress or representation options;
 - enabling persons affected to mandate a not-for-profit body to exercise their rights.
- provisions **reversing the burden of proof** for alleged plausible harm, where
 - the onus would be on the operator to prove they did all they reasonably could to assess and mitigate reasonably foreseeable negative impacts;
 - alternative redress and compensation mechanisms should be considered where no directly responsible party can be assigned.

⁵ These could be brought before the courts of the Member State where the operator is established or where the person habitually resides (unless the operator is a public authority of a Member State).

4. Ensuring Democratic Oversight

/ Risk or Impact Assessments

The duty to perform risk assessments or, preferably, **impact assessments**⁶ is essential to ensure the AIA's effectiveness. They aim to influence reiterative design processes to account for, and respond to harms, create transparency to enable learning, and hold relevant actors accountable.

To achieve these normative goals, the AIA should foresee due diligence obligations along the value chain by

- expanding the obligation to conduct a **fundamental rights impact assessment** foreseen in Art. 29 [EP] to both **providers and deployers** of foundation models and GPAI systems, within their respective sphere of influence.

Specifically:

- **Providers** must assess, monitor for, and mitigate reasonably foreseeable impacts resulting from development, training, and progressing insights about how their systems are deployed, including impact on rights and legitimate interests of third parties such as human dignity, non-discrimination, data protection, copyright, health, safety, workers' rights, the environment, and systemic risks to society (similar to Art. 34 DSA).
- **Deployers** must assess, monitor for, and mitigate reasonably foreseeable impacts resulting from the concrete deployment or integration into new services, including impact on rights and legitimate interests of third parties such as human dignity, non-discrimination, data protection, copyright, health, safety, workers' rights, the environment, and systemic risks to society (similar to Art. 34 DSA).
- In case of non-compliance with this requirement to conduct risk/impact assessments, **effective sanctions** must be foreseen.
- Risk/impact assessments must be **continuous** over the entire life cycle of a system, including its retirement phase, and **involve relevant stakeholders**, including interdisciplinary experts and representatives of affected people and marginalized communities. The regulator should issue more concrete **guidance** on risk/impact assessment procedures, documentation, and on how to enable participatory, inclusive, and interdisciplinary methods.
- To allow for public oversight, providers and at least high-risk deployers must be required to make risk/impact assessments, including fundamental rights and data protection impact assessments,⁷ **transparent** as part of Annex VIII registration duties.

⁶ While risk assessments identify and assess threats specific to the product or service under evaluation, impact assessments more broadly evaluate the implications for people, society, and the environment.

⁷ Including also information on data (in full respect of data protection); on data governance (Art. 10 AIA, including on data production process and internal or external parties' involvement); on technical standards (Art. 10, 11 AIA); information listed in recitals 60g and 60h [EP]; and documentation on periodic monitoring mechanisms.

- The AIA should include a provision guaranteeing **access to data** for researchers and civil society (comparable to Art. 40 DSA) and reporting obligations as well as a requirement to **conduct independent audits** of risk/impact assessments.

/ Standardization

The cornerstone of effective protection of people's rights and the interests of society are **strong, enforceable laws**. Recently proposed self-regulatory initiatives can be no alternative to setting rules on GPAI in the framework of democratic decision-making. Likewise, **standardization is limited** when it comes to preventing risks for fundamental rights as part of a highly technical product-safety approach. What is more, it has proven impossible to ensure that civil society and affected communities are adequately represented in the process, while industry is very strongly represented, adding to the potential democracy deficit.

In order to address these challenges, the AIA, with regard to GPAI:

- should further specify the **methodology for impact/risk assessments in a delegated act – not by technical standards** – and developed with the involvement of experts, civil society, and representatives of affected groups, including marginalized, or otherwise vulnerable, communities;
- can, however, for certain technical issues, mandate the Commission to develop **common specifications** (cf. Art. 41 AIA [COM]), ensuring effective inclusion of relevant stakeholders.

/ Research Exemption

The research exemption in Art. 2.5(d) [EP] should allow researchers to play their role as **innovators and critical observers**, while at the same time **prevent harm and abuse**.

To this end, the AIA must

- address the increasingly **unequal preconditions for public-interest research and commercial research and development** in this field;
- clarify the **beneficiaries and conditions** of a research exemption and ensure that any research activities are carried out in accordance with recognized **ethical and professionals standards** for scientific research;
- clarify that the research exemption also applies to testing GPAI and its impact on individuals and society **after its release**, including **open source** models;
- **prioritize delegated regulation** on the research exemption over, for example, voluntary codes of conduct;
- contain a provision on **research access to data** for academic and civil society researchers similar to Art. 40 DSA, covering, among others, GenAI systems (see above).

5. Complementing other Regulations

The AIA is just one piece of the regulatory landscape governing AI systems. It should be seen and designed in conjunction with other relevant legal frameworks, such as the ones listed below.

European lawmakers should:

- ensure **non-discrimination law** fully applies to the development and deployment of AI systems, including GPAI;
- explore the need to expand requirements in the **Digital Markets Act** to GPAI providers;
- clarify that to the extent that VLOPs and VLOSEs integrate GPAI, the **Digital Services Act's** obligation to monitor for and mitigate systemic risks also extends to risks that are the result of the development and deployment of GPAI on these platforms;
- ensure that **competition law** addresses the influence and market power of providers of foundation models and prevents anti-competitive practices;
- expand Art. 22 **General Data Protection Regulation** (GDPR) to apply to decisions generated by GPAI; and re-emphasize GDPR requirements in the AIA, including transparency about training data;
- ensure increased protection of **workers' rights** who contribute to developing, training, and fine-tuning GPAI systems, in particular through the **Platform Work Directive** and through the **Corporate Sustainability Due Diligence Directive**;
- consider the potential **environmental impact** of GPAI, inter alia through applying the Corporate Sustainability Due Diligence Directive to value chains of foundation models;
- ensure that the AIA does not preclude **sector specific regulation**. GPAI and GenAI will impact different sectors in specific ways, and existing regulations need to be evaluated and potentially updated accordingly.

Contact

AlgorithmWatch

Angela Müller, PhD (des.), Head of Policy & Advocacy

policy@algorithmwatch.org

<https://algorithmwatch.org>

University of Amsterdam – AI, Media & Democracy Lab

Natali Helberger, PhD, Distinguished University Professor Law & Digital Technology

N.Helberger@uva.nl

<https://www.uva-icds.net/ai-media-democracy/>