March 2022

RECOMMENDATIONS FOR A FUTURE-PROOF AI ACT

Contact
mark@futureoflife.org
+331616647630

CONTENTS

The Future of Life Institute (FLI) works to promote the benefits of technology and reduce their associated risks. FLI has become one of the world's leading voices on the governance of artificial intelligence (AI) and created one of the earliest and most influential sets of governance principles, the Asilomar AI Principles. FLI maintains a large network among the world's top AI researchers in academia, civil society, and private industry.

Since progress in AI can be very rapid, it is particularly important to ensure that the AI Act is prepared for these technological changes.

GENERAL PURPOSE SYSTEMS

NEW ARTICLE

Article 3 — 'General purpose AI system' means an AI system that is able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation, etc, and is able to have multiple intended and unintended purposes.

JUSTIFICATION

An image recognition system that identifies signs of skin cancer has an intended purpose. Another system that identifies potholes in roads from images also has an intended purpose. A system able to identify skin cancer and potholes, has two (quite different) intended purposes. General purpose AI systems can have many more intended purposes as well as many unintended uses. The definition of general purpose AI systems hinges on its ability to do several different tasks.

General purpose AI systems are software, which means they can very quickly be applied to a wide range of areas - much faster than the EU can adopt new acts. Therefore, the solution is to cover them in this regulation by default, and to ensure the responsibility for their safety is not just on EU companies, but shared with the creators of general purpose AI systems. MEP Voss (EPP, DE) suggested a version of this recommendation in his first draft opinion on the AI Act for the Legal Affairs (JURI) Committee.

NEW ARTICLE

Article 4a: Obligations of providers of general purpose AI systems

Providers of general purpose AI systems shall:

a) ensure that their general purpose AI systems are compliant with the requirements set out in Article 15 in Chapter 2;
b) comply with the other requirements set out in Chapter 2 to the fullest extent possible;
c) assess the reasonably foreseeable misuse of their systems;
d) provide instructions and information about the safety of these systems to users and other relevant stakeholders in the supply chain;
e) regularly assess whether the AI systems have presented any new risks, including risks discovered when investigating novel use cases;
f) register their systems in the EU database referred to in Article 60.

JUSTIFICATION

General purpose AI systems are are are able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation, etc. These systems are trained on broad data that can be adapted to a wide range of downstream tasks and applications, and have multiple intended and unintended purposes.

The wide range of applications for which general purpose systems can be used means that any flaw can have overarching effects on many sectors – there is a single failure mode that can affect many downstream AI applications. Researchers recently showed that one general purpose AI system, for example, had an anti-Muslim bias. If left unaddressed, a bias of this type could affect media articles, educational materials, chatbots, and other uses that will only be discovered when SMEs experiment with these systems. The potential use of general purpose AI systems for many tasks with different levels of risk justifies pre-market

and post-market obligations.

Furthermore, general purpose AI systems will be integrated into various services and products. (European) companies integrating (mainly American) general purpose systems will be unable to understand the full risks involved. Product safety regulation commonly uses the concept of 'reasonably foreseeable use.' It is reasonable to ask providers to try to foresee the potential misuses of their general purpose AI systems because of the likely significant economic and societal impacts of these systems. They should address these risks to health, safety and fundamental rights beforehand, while acknowledging that not all such uses can be foreseen. Developers of these systems are best placed to estimate foreseeable uses and misuses.

NEW ARTICLE

Article 4b: Conformity assessment for general purpose AI systems
Providers of general purpose AI systems shall ensure that their systems undergo a conformity assessment prior to their placing on the market or putting into service. The conformity assessment procedure will be based on assessment of the quality management system and assessment of the technical documentation, with the involvement of a notified body, referred to in Annex VII. Where the compliance of the AI systems with the requirements set out in Chapter 2 of this Title has been demonstrated following that conformity assessment, the providers shall draw up an EU declaration of conformity in accordance with Article 48 and affix the CE marking of conformity in accordance with Article 49.

JUSTIFICATION

As general purpose AI systems will be integrated into many other AI applications, requiring providers to undergo a conformity assessment will create lower regulatory burdens for the hundreds of potential companies using those systems as foundations to their more narrow applications.

As ALLAI notes, general purpose systems ought to be included in scope to avoid a situation where the burden of bringing these systems in compliance with the AIA falls entirely on 'downstream' users of the general purpose AI systems. Downstream users would otherwise be the ones that have to bring the systems in line with the requirements for high risk AI, which might be too much of a burden, especially for SME's and micro enterprises, or perhaps even prove to be technically impossible. Even if the general purpose AI developer would help downstream users with the technicalities of complying with the AIA, it places the latter in a fully dependent position, without having the appropriate means to seek redress when the general purpose AI system causes damage.

A third-party conformity assessment will guarantee to European users that these systems are accurate, robust, withstand cyberattacks, and can be trusted for use.

NEW ARTICLE

Article 4c: Conditions for other persons to be subject to the obligations of a provider
Any person who places on the market, puts into service or uses a general purpose AI system in any of the circumstances listed in Article 28 shall be considered one of the providers of the system subject to the provisions of this Regulation. The developer of the general purpose AI system will be considered the provider unless the Article 28 conditions apply.

JUSTIFICATION

While various stakeholders in the supply chain should have specific responsibilities, the original creators of these systems (almost exclusively large multinational companies head-

quartered in the US or China) should be treated as providers because they are best suited to conduct assessments of their systems. This will create lower regulatory burdens for the hundreds of potential companies using those systems as foundations to their more narrow applications.

When general purpose AI systems are used in the AI value chain, it should be possible for multiple companies to be considered providers under the regulation: the initial builder of the general-purpose system and the company or companies adapting it downstream.

MANIPULATION

TEXT PROPOSED BY THE COMMISSION	SUGGESTION
Article 5(1): 1. The following artificial intelligence practices shall be prohibited: (a) the placing on the market, putting into service or use of an AI system ~~that deploys subliminal techniques beyond a person's consciousness in order to materially distort~~ a person's behaviour ~~in a manner that~~ causes or is likely to cause that person or another person ~~physical or psychological harm~~; (b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons ~~due to their age, physical or mental disability, in order to~~ materially distort the behaviour of a person pertaining to that group ~~in a manner that causes or is likely~~ to cause that person or another person ~~physical or psychological harm~~;	Article 5(1): 1. The following artificial intelligence practices shall be prohibited: (a) the placing on the market, putting into service or use of an AI system which significantly materially distorts a person's behaviour or causes or is likely to cause that person, another person or society at large significant harm; (b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons by significantly materially distorting the behaviour of a person pertaining to that group to cause that person, another person or society at large significant harm;

JUSTIFICATION

We think that the proposed addition of 'significant' to Article 5 by MEP Voss (EPP, DE) in his first draft opinion on the AI Act for the Legal Affairs (JURI) Committee is helpful, because it captures the most concerning systems whilst preventing the risk of regulatory overreach. However, we recommend avoiding the addition of 'with the objective to' as it can make the manipulation protection meaningless because AI systems are arguably never employed to purposefully distort or cause harm. For example, internal Facebook research from 2018 shows that 64% of extremist group joins are because of automated recommendations - which were obviously never designed for this purpose.

Some recent AI systems have also not caused much harm to individuals but have harmed society at large. The Cambridge Analytica micro-targeting of voters in the Brexit referendum is a good example and we would expect these kinds of incidents to proliferate beyond the boundaries of social media.

Other upcoming AI regulations like the Digital Services Act also include some protections against societal harms but cover only very large online platforms, whilst the AI Act should also protect against societal harms from smaller companies. The new upcoming political advertising regulation may be able to protect against societal harm arising from political advertising, but it will not apply to other societal harms caused or likely to be caused by AI. In a paper about societal harm, researcher Nathalie Smuha provided several other examples, including harms to rule of law caused by AI systems used in the context of law enforcement, public administration or the judicial system.

FLI therefore recommends prohibiting AI manipulation that causes significant harm to society as well as individuals.

REGULATORY FLEXIBILITY

TEXT PROPOSED BY THE COMMISSION	SUGGESTION
Article 7(1): 1. The Commission is empowered to adopt delegated acts in accordance with Article 73 to update the list in Annex III by adding high-risk AI systems where ~~both~~ of the following conditions ~~are~~ fulfilled:	Article 7(1): 1. The Commission is empowered to adopt delegated acts in accordance with Article 73 to update the list in Annex III by adding high-risk AI systems where one of the following conditions is fulfilled:

JUSTIFICATION

The high-risk requirements in the AI Act apply to only eight sectors and the proposal provides no means of adding additional sectors. In the future, AI applications may pose a risk to human rights, safety and health in an unforeseen sector. For example, the use of AI in finance, environment, healthcare and many other areas, which are currently not covered in the proposal, could become high-risk. Together with the European Consumer Organisation BEUC, FLI believes that regulatory flexibility needs to be expanded to ensure a future-proof response to AI developments in various areas of life.

WHISTLEBLOWERS

NEW RECITAL

Recital 48a:

In order to protect the developers of AI systems against retaliation from their employers and colleagues, and to prevent misconduct or breaches of laws and regulations, developers should be able to rely on EU whistleblower protections as set by Directive (EU) 2019/1937 of the European Parliament and of the Council.

JUSTIFICATION:

As AI applications grow ever more complex, it will become increasingly difficult to know whether an application constitutes a health risk or a potential violation of human rights. In fact, some of the only people who will be able to determine the nature and extent of risks and harms will be the developers themselves. Therefore, these developers should be provided with the right to voice concerns to a relevant supervisory authority through a dedicated channel if internal company channels are insufficient and should be able to rely on EU whistleblower protections (Directive (EU) 2019/1937). Recent controversies surrounding the employment of experts in AI ethics by major private companies make it apparent that such whistleblower protections may be necessary for industry experts to feel comfortable about raising concerns to outside authorities.

CONFLICTS OF INTEREST

TEXT PROPOSED BY THE COMMISSION	SUGGESTION
Article 52(1): 1. Providers shall ensure that ~~AI systems intended to interact with natural persons are designed and developed in such a way that~~ natural persons are informed that ~~they are interacting with an AI system~~, unless this is obvious from the circumstances and the context of use. 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 52(1): 1. Providers shall ensure that natural persons are informed that an AI system is involved in the product or service they are using, unless this is obvious from the circumstances and the context of use. Providers shall explicitly state possible conflicts of interest of the AI system. 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.

JUSTIFICATION

Consumers can be harmed when an AI provider enters into an undisclosed business arrangement and makes recommendations based on the interests of the third party instead of the interests of the consumer. Legal protections against conflicts of interest will become increasingly important as more capable AI systems take over high-stakes medical, legal or financial services. At a minimum, FLI (alongside AlgorithmWatch) believes that greater transparency for AI systems about their purpose and logic should be required.

SANDBOXES

TEXT PROPOSED BY THE COMMISSION	SUGGESTION
Article 53(1): 1. AI regulatory sandboxes established by one or more Member States competent authorities or the European Data Protection Supervisor shall provide a controlled environment that 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. This shall take place under the direct supervision and guidance by the competent authorities with a view to ensuring compliance with the requirements of this Regulation and, where relevant, other Union and Member States legislation supervised within the sandbox.	Article 53(1): 1. AI regulatory sandboxes established by one or more Member States competent authorities or the European Data Protection Supervisor shall provide a controlled environment that 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. This shall take place under the direct supervision and guidance by the competent authorities with a view to ensuring compliance with the requirements of this Regulation and, where relevant, other Union and Member States legislation supervised within the sandbox. Participants in the AI regulatory sandboxes, in particular small-scale providers, are granted access to pre-market services, such as legal support, preliminary registration of their AI system, R&D support services, and to all the other relevant elements of the Union's AI ecosystem and other Digital Single Market initiatives. SMEs outside the Union can apply for participation in the AI regulatory sandboxes.

JUSTIFICATION

Enhanced sandboxes could offset some of the regulatory burden introduced through the Act by offering additional services to participating businesses, such as legal support, lab-to-market insurance and fiscal incentives for R&D activities. Moreover, the EU should consider opening access to sandboxes to SME's from outside the Union. This would promote the dissemination of EU standards globally. The EU could also facilitate input from AI experts in civil society and academia through the sandboxes to help ensure that the guidance provided to businesses remains state-of-the-art. The Committee on Industry, Research and Energy (ITRE) suggested a version of this recommendation in their draft opinion of the AI Act.

TEXT PROPOSED BY THE COMMISSION	SUGGESTION
Article 53(6): 6. The modalities and the conditions of the operation of the AI regulatory sandboxes, including the eligibility criteria and the procedure for the application, selection, participation and exiting from the sandbox, and the rights and obligations of the participants shall be set out in implementing acts. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 74(2).	Article 53(6): 6. The modalities and the conditions of the operation of the AI regulatory sandboxes, including the eligibility criteria and the procedure for the application, selection, participation and exiting from the sandbox, and the rights and obligations of the participants shall be set out in implementing acts. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 74(2). The Commission shall establish the EU AI Regulatory Sandbox. The Commission shall coordinate the procedures and activities of the EU AI Regulatory Sandbox with national and local authorities.

JUSTIFICATION

If Europe truly wants to be at the forefront of AI development, its approach to sandboxes could be made more ambitious. FLI foresees a significant risk that companies will be presented with a bewildering range of sandbox schemes in different Member States. This disparity risks fracturing the Single Market and slowing down AI development. A stronger approach would involve the establishment of a pan-European sandbox, which would be accessible through a single online AI portal: an EU AI Regulatory Sandbox. The Committee on Industry, Research and Energy (ITRE) suggested a version of this recommendation in their draft opinion on the AI Act.

AI BOARD

TEXT PROPOSED BY THE COMMISSION	SUGGESTION
Article 56(2): 2. The Board shall provide advice and assistance to the Commission in order to: (a) contribute to the effective cooperation of the national supervisory authorities and the Commission with regard to matters covered by this Regulation; (b) coordinate and contribute to guidance and analysis by the Commission and the national supervisory authorities and other competent authorities on emerging issues across the internal market with regard to matters covered by this Regulation; (c) assist the national supervisory authorities and the Commission in ensuring the consistent application of this Regulation.	[...] (d) Monitor the latest technological developments, impacts on start ups and SMEs, and the effectiveness of existing regulations for general purpose AI systems.

JUSTIFICATION

The nature of general purpose AI systems means that they can be applied to countless different kinds of applications in unexpected ways. When the GPT-3 language model grew beyond a certain scale, for example, researchers discovered that it could not just generate text but also perform calculations. This was an unexpected feature, and future (even larger) general purpose systems will likely to be able to perform other unexpected and potentially risky tasks. The AI Board should monitor these developments and be able to recommend changes. The Committee on Industry, Research and Energy (ITRE) suggested a version of this monitoring recommendation in their draft opinion of the AI Act.

TEXT PROPOSED BY THE COMMISSION	SUGGESTION
Article 58: When providing advice and assistance to the Commission in the context of Article 56(2), the Board shall in particular: (a) collect and share expertise and best practices among Member States; (b) contribute to uniform administrative practices in the Member States, including for the functioning of regulatory sandboxes referred to in Article 53; (c) issue opinions, recommendations or written contributions on matters related to the implementation of this Regulation, in particular (i) on technical specifications or existing standards regarding the requirements set out in Title III, Chapter 2, (ii) on the use of harmonised standards or common specifications referred to in Articles 40 and 41, (iii) on the preparation of guidance documents, including the guidelines concerning the setting of administrative fines referred to in Article 71.	[...] (iv): on the classification rules for high-risk AI systems and amending the list of high-risk areas set out in Chapter 1.

JUSTIFICATION

The high-risk requirements in the AI Act apply to only eight sectors and the proposal provides no means of adding additional sectors. In the future, AI applications may pose a risk to human rights, safety and health in an unforeseen sector. For example, the use of AI in finance, environment, healthcare and many other areas, which are currently not covered in the proposal, could become high-risk. Together with the European Consumer Organisation BEUC, we believe regulatory flexibility needs to be expanded to ensure a future-proof response to AI developments in various areas of life.

PUBLIC SECTOR CAPACITY

TEXT PROPOSED BY THE COMMISSION	SUGGESTION
Article 59(2): 2. Each Member State shall designate a national supervisory authority among the national competent authorities. The national supervisory authority shall act as notifying authority and market surveillance authority unless a Member State has organisational and administrative reasons to designate more than one authority.	Article 59(2): 2. Each Member State shall designate a national supervisory authority among the national competent authorities with sufficient regulatory capacity. The national supervisory authority shall act as notifying authority and market surveillance authority unless a Member State has organisational and administrative reasons to designate more than one authority.

JUSTIFICATION

AI is a major potential source for economic growth that must be facilitated by civil servants who understand the latest technological developments. Governments should therefore view increased public sector capacity for AI development as an opportunity, rather than as a burden. Currently, the Commission estimates that the implementation of the proposal will require no more than 25 Full Time Equivalent civil servants per Member State. This estimate is likely to underestimate the transformative impact that AI will have on both societies and their public sectors. Beyond the minimal bar set by the Commission, both EU institutions and Member States should consider investing extra resources in public sector capacity in order to i) ensure companies are able to get quick answers from regulators on whether a sandbox application can be placed on the market; ii) improve understanding of where public research funds can best be directed, iii) quickly publish non-personal data from local, regional and national authorities to improve public services.

EU DATABASE

TEXT PROPOSED BY THE COMMISSION	SUGGESTION
Article 62(1): 1. Providers of high-risk AI systems placed on the Union market shall report any serious incident or any malfunctioning of those systems which constitutes a breach of obligations under Union law intended to protect fundamental rights to the market surveillance authorities of the Member States where that incident or breach occurred.	Article 62(1): 1. Providers of high-risk AI systems placed on the Union market shall report any serious incident or any malfunctioning of those systems which constitutes a breach of obligations under Union law intended to protect fundamental rights to the market surveillance authorities of the Member States where that incident or breach occurred. The market surveillance authorities of the Member States shall report these incidents to the EU database referred to in Article 60.

JUSTIFICATION

The development of AI systems is happening at breakneck speed and their safety implications often only become known after they are placed on the market. The Boeing 737 MAX, for example, had been tested for many years and was certified by the U.S. Federal Aviation Administration in March 2017. It took two years and two plane crashes before investigators discovered that an AI-based software system within the cockpit, the Manoeuvring Characteristics Augmentation System (MCAS), produced fatal nose-down commands without an override option for pilots.

In our view, AI advancement would benefit from a clear overview of safety incidents at the European level, because doing so will make it easier to analyse what research or regulation may be necessary as trends emerge across the Single Market. Therefore, and in the spirit of the existing Seveso directive on industrial accidents, FLI recommends that Member States also report safety incidents to the EU database.

RIGHT OF COMPLAINT

NEW ARTICLE

Article 68a: Right to lodge a complaint with a supervisory authority

1. Every citizen who considers that his or her right to protection of health, safety and fundamental rights has been infringed by the use of a prohibited AI system or a high-risk AI system shall have the right to lodge a complaint with the authority in charge to handle complaints in the Member State of his or her habitual residence, place of work or place of the alleged infringement.
2. The supervisory authority with which the complaint has been lodged shall investigate the subject matter of the complaint and inform the complainant on the progress and the outcome of the investigation within a reasonable time period.

JUSTIFICATION

This right of appeal (Article 45) is an important safeguard of fundamental rights. It will ensure that, for example, a trade union can appeal the approval of systems that use facial recognition in ways not intended or foreseen by the notified body. However, the right to appeal decisions of notified bodies alone provides insufficient protection of fundamental rights, because the proposal allows for many (high-risk) AI applications to be put on the market after self-assessment and without third-party involvement.

Under the current draft, the relevant national supervisory authority is the only body that can act when a provider of (high-risk) AI systems overlooks or evades a legal requirement. The implications of this are stark. If someone falls victim to an AI system that, for example, "deploys (harmful) subliminal techniques beyond a person's consciousness", then they may not be able to file a complaint through a dedicated process. FLI therefore proposes the inclusion of an 'individual right to lodge a complaint', partially inspired by the existing remedy under the GDPR. The Committee on Legal Affairs (JURI) has also suggested a version of this recommendation in their draft opinion of the AI Act.