



EUROPE | NOVEMBER 2022

Artificial Intelligence Act:

A Policy Prototyping Experiment

*Operationalizing the Requirements
for AI Systems – Part I*



E

P

N

L



R

NORBERTO NUNO
GOMES DE ANDRADE

ANTONELLA ZARRA



About Open Loop

Open Loop is a global program that connects policymakers and technology companies to help develop effective and evidence-based policies around AI and other emerging technologies.

The program, supported by Meta (previously Facebook), builds on the collaboration and contributions of a consortium composed of regulators, governments, tech businesses, academics, and civil society representatives. Through experimental governance methods, Open Loop members co-create policy prototypes and test new and different approaches to laws and regulations before they are enacted, improving the quality of rulemaking processes in the field of tech policy.

This report presents the findings and recommendations of the first part of the Open Loop's policy prototyping program on the European Artificial Intelligence Act, which was rolled out in Europe from June 2022 to July 2022 and in partnership with Estonia's Ministries of Economic Affairs and Communications and Justice and the Malta Digital Innovation Authority (MDIA).

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Cite this report

Andrade, Norberto Nuno Gomes de, and Antonella Zarra. "Artificial Intelligence Act: A Policy Prototyping Experiment: Operationalizing the Requirements for AI Systems – Part I" (2022), at https://openloop.org/reports/2022/11/Artificial_Intelligence_Act_A_Policy_Prototyping_Experiment_Operationalizing_Reqs_Part1.pdf

Acknowledgements

This policy prototyping program was co-designed and facilitated by **Meta** in collaboration with our partners from the **Government of Estonia** and the **Malta Digital Innovation Authority (MDIA)**.

We want to thank in particular Henrik Trasberg, Legal Advisor on AI and New Technologies, Estonian Ministry of Justice; Ott Velsberg, Government Chief Data Officer, Estonian Ministry of Economic Affairs and Communications; Kenneth Brincat, CEO, MDIA and Ian Gauci and Gordon Pace, Advisors, MDIA.

A special thanks to Prof. Bart Schermer and Jord Goudsmit from **Considerati** and Volker Bilgram and Rebekka Bogner from **Hyve Innovate** for their invaluable contribution to this project.

We are particularly thankful to all the speakers and participants of the Open Loop Program on the EU AI Act kickoff event for informing this report through their excellent contributions: Dragoş Tudorache, Member of the **European Parliament** (MEP) and co-rapporteur of the AI Act for his opening keynote; the panelists Evi Fuelle, Global Policy Director at **Credo AI**, Horst Heitz, Chair of the European Steering Board at **SME Connect**, and Karine Perset, Head of AI Unit and OECD.AI at **OECD**; and more than 80 attendees that participated in the event.



We would like to thank the following companies for their partnership and participation. Without their commitment and active involvement this project would not have been possible:

2021.ai Denmark
ASC27 Italy
Armillar AI Canada
Ask Mona France
AUNOA Spain
Bluetensor Srl Italy
Credo AI United States
Deeploy Netherlands
DLabs.AI Poland
Elara Tech Georgia
Enzai Technologies Limited Ireland
Evo Italy/United Kingdom
Flawless United Kingdom
Flytxt BV Netherlands
Future Intelligence Greece
Gossik AG Switzerland
indigo.ai Italy
Invrision Italy
iov42 United Kingdom/Austria
Irida Labs Greece
Keeper Data Tech Spain
Keyless United Kingdom/Italy/Singapore
Kodex AI Germany
LearnerShape United Kingdom
M3i Industry-in-Clinic-Platform Germany
Manent.AI Italy
ML cube Italy
Myco Italy

N Robotics Germany
NAIX Technology Germany
Palko Portugal
Peregrine Technologies GmbH Germany
Polaris Engineering S.p.A. Italy
Qubit Ventures Germany
super.AI Germany
SynerScope bv Netherlands
TAWNY Germany
Telesoftas Lithuania
Tgs Baltic Lithuania/Estonia/Latvia
The Newsroom Portugal
The VR Hive United Kingdom
Translated Italy
Travers Smith United Kingdom
Vedrai SPA Italy
Virtuleap Portugal/United States/
United Kingdom/Spain
Visor.ai Portugal
Vixtape Portugal
Voxist France
Vytautas Magnus University Lithuania
Watermelon Netherlands
Wayve United Kingdom
WingDriver United States
yData Labs Inc Portugal
Zupervise United Kingdom/India



Thank you in particular to the individual experts that represented the participating companies throughout the program:

Rasmus Hauch, Björn Preuss, Colum Donnelly, Nicola Grandis, Marion Carré, Fernando Perez, Jonni Malacarne, Ehrik Aldana, Susannah Shattuck, Evi Fuelle, Bastiaan van de Rakt, Maciej Karpicz, Shemmy Majewski, Negar Vahid, Ryan Donnelly, Kaitlin Goodrich, Stuart Lawrence, Prateek Kapadia, Anargyros Sideris, Benno Staub, Gianluca Maruzzella, Fabrizio Dini, Alexis de Vienne, Thomas Charisis, Christos Theocharatos, Ezequiel Paura, Pierluigi Failla, Claus Lang, Maury Shenk, Florian Neumeier, Simone Gabriellini, Alessandro Nuara, Francesco Trovò, Tommaso Demattè, Davide Fanale, Nicola Caporaso, Elisa Czerski, Ramin Karbalaie, Rui Dias Ferreira, Philip Meier, Alessandro Lazzeri, Federico Cesari, Hakki Ercosman, Sina Youn, Jan-Kees Buenen, Marco Maier, Bart Kappel, Mindaugas Civilka, Jenny Romano, Lorenzo Mora, Pedro Henriques, James Black, Sébastien Bratières, Shawn Curran, Hossein (Kian) Sarpanah, Amir Bozorgzadeh, Victòria Brugada-Ramentol, Bebiana Moura, Gonçalo Consiglieri, Michael Fiorentino, Karel Bourgois, Tomas Krilavičius, Darius Amilevičius, Alexander Wijninga, Sarah Gates, Daniel Quirke, André Azevedo, Fabiana Clemente, Janhvi Pradhan Deshmukh, Philip Dawson



We would also like to thank the Open Loop's partners and observers from business, academia, civil society, international institutions, and governmental authorities:

Members of European Parliament

- **Eva Maydell**
ITRE Rapporteur of the AI Act (BG, EPP)
- **Ivan Štefanec**
President of SME Europe (SK, EPP)

International Institutions / Governmental Authorities

- **Alessandro Fusacchia**
Member of the Italian Parliament, Co-ordinator of Parliamentary Intergroup on AI
- **Karine Perset**
Head of AI Unit and OECD.AI, Organisation for Economic Co-operation and Development (OECD)
- **Andras Hlacs**
AI Policy Analyst, OECD.AI, Organisation for Economic Co-operation and Development (OECD)
- **Luca Carabetta**
Member of the Italian Parliament
- **Işıl Selen Denemeç**
Head of Legal Department, Digital Transformation Office of the Presidency of the Republic of Türkiye
- **Richard Nevinson**
Head of Digital Economy, Information Commissioner's Office (ICO)
- **Zümrüt Müftüoğlu**
Expert, Digital Transformation Office of the Presidency of the Republic of Türkiye

Academia / Think Tanks

- **Andrea Bertolini**
EURA Centre, Scuola Superiore Sant'Anna Pisa
- **Johann Laux**
Oxford Internet Institute
- **David Osimo**
Lisbon Council
- **Joshua Ellul**
University of Malta and Former chairman, Malta Digital Innovation Authority
- **Eduard Fosch**
Leiden University
- **Klaus Heine**
DIGOV Centre Rotterdam University
- **Evert Stamhuis**
DIGOV Centre Rotterdam University
- **Nicolaos Voros**
University of the Peloponnese
- **Fabiana di Porto**
University of Salento and LUISS University
- **Risto Uuk**
The Future of Life Institute
- **Giovanni Sartor**
University of Bologna and European University Institute
- **Virginia Dignum**
Umeå University

Industry Associations



Civil Society



Artificial Intelligence Act: A Policy Prototyping Experiment

Overview of the Open Loop Program on the AI Act and the stakeholders involved



Executive summary	10
Introduction	14
Research goals and methodology	16
The Open Loop Forum.....	18
Scope and limitations.....	23
Activity 1: Taxonomy of AI actors	24
Task 1: Determining your role under the AIA.....	25
Task 2: User and provider.....	27
Task 3: Intended purpose of the system.....	27
Task 4: Scenarios on actors.....	28
Observations.....	33
Activity 2: Managing risk	34
Task 1: Risk assessment appetite.....	36
Task 2: Understanding known and foreseeable risks.....	36
Task 3: Confidence in assessing risks and misuse of AI systems.....	37
Task 4: Scenarios on risks.....	39
Observations.....	41
Activity 3: Setting up data quality requirements	42
Task 1: Feasibility of data requirements.....	43
Task 2: European Parliament amendments.....	45
Observations.....	46

Activity 4: Technical documentation **47**

Task 1: Describing system characteristics48

Task 2: Level of detail of description 51

Task 3: Completeness of description52

Task 4: Information about functioning and control.....53

Discussion.....54

Observations.....56

Activity 5: Assuring transparency and human oversight **58**

Task 1: Designing for transparency59

Task 2: Human oversight..... 61

Discussing AI systems’ design, interpretation, use, and risk management62

Observations.....64

Bonus activity: Regulatory sandboxes **65**

Observations.....70

Conclusion and recommendations **71**

Recommendations73

Endnotes **74**

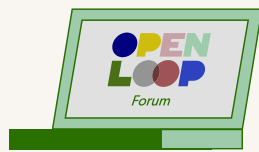


Executive summary

Executive summary

In this first phase of the Open Loop policy prototyping program on the EU AI Act (AIA), we tested selected articles of the draft proposal to assess, in practice, how understandable, (technically) feasible, and effective they are.

The program adopts a policy prototyping approach and is structured into three distinct, but connected, phases, namely i) operationalizing the requirements for AI systems, ii) regulatory sandbox, and iii) taxonomy of AI actors. The program deploys qualitative and quantitative methods and involves industry partners, EU institutions, governmental entities, regulatory authorities, academics, and other nongovernmental organizations.¹ Policy prototyping is a methodology to assess the efficacy of a policy by testing it in a controlled environment first. Policy prototyping applies a design thinking approach, which is common in product and service design, to the development of law and policy.



For this first phase of the project, we enlisted 53 AI companies to participate in the Open Loop Forum (OLF),

a dedicated online platform where they met to discuss topics and complete several research-related tasks. Over a period of three weeks (from June 2022 to July 2022), participants were invited to provide their feedback and views on selected articles of the AIA. A combination of multiple-choice and open-ended questions based on five themes from the AIA:

- **Taxonomy of AI actors (Article 3)**
- **Risk management (Article 9)**
- **Data quality requirements (Article 10)**
- **Technical documentation (Article 11)**
- **Transparency and human oversight (Articles 13 and 14)**

We also added a "bonus activity" for those interested in regulatory sandboxes. The overall picture (based on the provisions from the AIA that we presented to the participants) was that for the majority of the participants the provisions in the AIA were clear and feasible and could contribute to one of the goals of the legislator: to build and deploy trustworthy AI. However, there were several areas in the AIA with room for improvement and some provisions that might even hinder the other goal of the legislator: enabling the uptake of AI in Europe.



Activity 1: Diving into the AI ecosystem – taxonomy of actors (Article 3 AIA)

The taxonomy of AI actors (providers, users, etc.) was mostly clear for the participants. However, while the definitions were clear "on paper," the participants did point out that in reality the roles of users and providers are not as distinct as the AIA presupposes, particularly in the context of the dynamic and intertwined relationships and practices that occur between the various actors involved in the development, deployment, and monitoring of AI systems. This raises questions as to who should be held responsible for the requirements in the AIA and who is responsible when these requirements are not met. We may tentatively conclude that the approach of the AIA – where there is a producer providing products to a user – does not fully match the increasing complexity and granularity of the reality of the AI ecosystem.



Activity 2: Managing risk (Article 9 AIA)

In the second activity, we assessed the willingness of the participants to manage risks even when they are not classified as high risk in the AIA. We also tested how confident the participants were that they could meet all the mentioned criteria (e.g., whether they could assess and predict risks and misuse). Most participants responded that they would still perform a risk assessment even if their AI systems are not high risk. They also seemed to understand what was meant by "known and foreseeable risk" and

were generally convinced that they could do this type of risk assessment and prediction for the intended purposes of their AI systems. In fact, participants pointed out that it was difficult for them to predict and anticipate how users or third parties would use their AI systems. Interestingly, participants seemed to focus more on the cause of risks (e.g., model drift and biased data) and less on the impact of these risks on natural persons (e.g., reputational damage, exclusion and discrimination).



Activity 3: Setting up data quality requirements (Article 10 AIA)

Participants seemed to agree that while the data requirements listed in the AIA cover areas that are relevant to consider when developing and deploying AI systems, the absolute nature of how these requirements are phrased and how they should be met (completeness, free of errors, etc.) is highly unrealistic to achieve. The "best effort" approach that was introduced by the European Parliament (i.e., ensuring a data set is free of errors and complete to the best extent possible), is seen as an improvement. Furthermore, participants underlined the importance of receiving guidance on the operationalization of these requirements. They suggested defining and documenting a series of practical guidelines to make complying with the data quality requirements more feasible.



Activity 4: Drawing up the technical documentation (Article 11 AIA)

The activity on technical documentation was aimed at assessing whether the participants are able to draw up the technical documentation for their AI systems as per Article 11 of the AIA. The responses show that there is a level of uncertainty and an operational gap on how participants should comply with some of the requirements listed in the draft proposal, specifically in terms of documenting the characteristics of their AI systems. This makes us tentatively conclude that the high degree of prescriptiveness of the AIA proposal may curtail the level of discretion needed to fulfill its requirements. In fact,

by listing a multitude of specific requirements, highly prescriptive laws such as the AIA often end up also requiring additional prescriptive guidance, which can make them more difficult to comply with, as there is less flexibility. While the AIA improves legal certainty by making it more explicit what is expected of providers, it unintendedly poses additional challenges to AI companies when it comes to interpreting and complying with such legal requirements. This contrasts with non-prescriptive laws that have a high level of abstraction, where more is left to interpretation in practice (e.g., through the guidance of the regulator, creation of market standards, and/or jurisprudence). Given the high level of detail in the AIA, further guidance by the legislator or the regulator on describing their AI systems is desired by the participants.



Activity 5: Assuring transparency and human oversight (Articles 13 and 14 AIA)

When it comes to transparency and human oversight, most participants seem to argue that AI systems should be designed in a way that there is no or little need for technical skills to use them. Like the AIA, participants distinguish between the operation of an AI system and the oversight of that system. The latter requires a different level of skills, which implies that different types of information, explanations, and instructions are needed for different target groups. The participants foresee challenges when it comes to providing transparency and are also unsure how they should balance explainability and model performance. From this activity, we may conclude that the AIA would benefit from clarifications on the way in which different target audiences² should be informed about the operation of the AI system.



Bonus activity: Regulatory sandboxes (Article 53 AIA)

What we can learn from the reactions of the participants is that the regulatory sandbox provision in the AIA can be an important mechanism not only to foster innovation, but also to strengthen compliance. However, there are several prerequisites that must be met for a sandbox to be

effective, such as legal certainty and a collaborative environment. These are important elements that are not covered (in depth) in Article 53 of

the proposed regulation. These elements could be addressed through implementing acts and guidance at the sandbox level.

Recommendations

While the scope of this exercise was limited and more research is needed, we tentatively formulate the following recommendations for improving the AIA and its effectiveness in reaching the goals of the legislator:



Consider **revising/expanding the taxonomy of AI actors** in Article 3 and/or more accurately describe possible interactions between actors (e.g., co-production of AI systems and use of open-source tooling) to more accurately reflect the AI ecosystem.



Given the difficulty in assessing "reasonably foreseeable misuse" (Article 9) and the limited focus on the impact of risks, **provide guidance on risks and risk assessment**, in particular for startups and SMEs.



Provide more concrete guidance, methodologies, and/or metrics for assessing the data quality requirements through, e.g., subordinate legislation and/or soft law instruments, standardization, or guidance from the regulator (Article 10).



Revise the data quality requirements "error free" and "complete" as they are considered unrealistic and unfeasible (Article 10).



Provide more concrete guidance, templates, and/or metrics for the technical documentation through, e.g., subordinate legislation and/or soft law instruments, standardization, or guidance from the regulator (Article 11).



Avoid a situation where the requirement for technical documentation becomes a "paper tiger" by ensuring **sufficient and sufficiently qualified staff to actually assess the technical documentation** (Article 11).



Consider **distinguishing more clearly between different audiences** for explanations and other transparency requirements (Articles 13 and 14) in the AIA.



The AIA's success hinges on the ability to execute and enforce the regulation. Therefore, it is important to ensure that the future workforce contains **enough qualified workers**, in particular when it comes to **human oversight of AI** (Article 14).



Maximize the potential of regulatory sandboxes to foster innovation, strengthen compliance, and improve regulation. Ensure that, through implementing acts and guidance, **conditions for effective AI regulatory sandboxes are created** (e.g., collaboration, transparency, guidance and legal certainty, and protection from enforcement) (Article 53).



Introduction

EU AI Act and policy prototyping

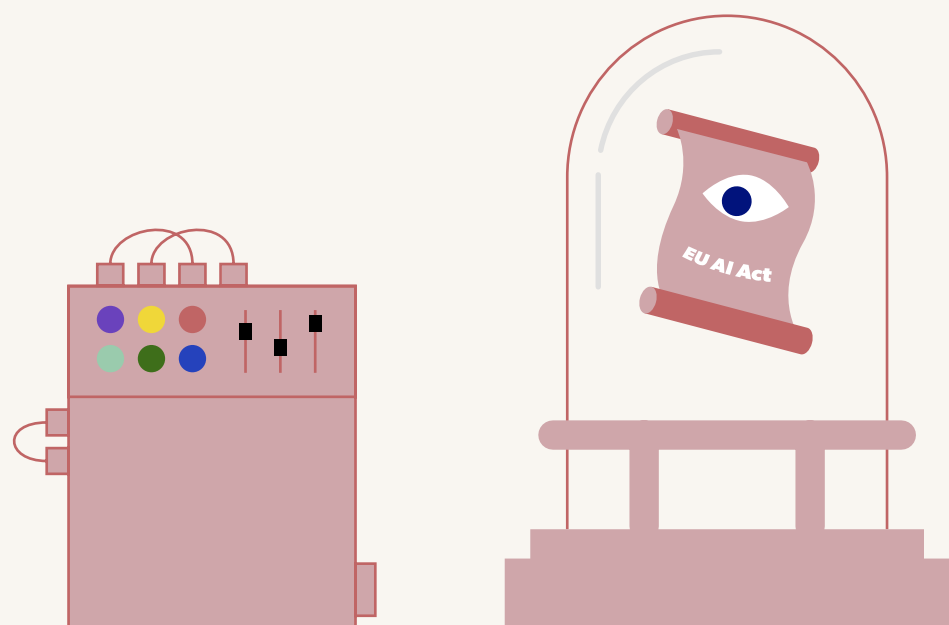
On April 21, 2021, the European Commission published the first draft of the EU AI Act (AIA).³ The Act has been debated by researchers, experts, AI companies, civil society, Member States, and many others. In this Open Loop policy prototyping program, we tested how clear and technically feasible selected articles of the AIA are in practice.

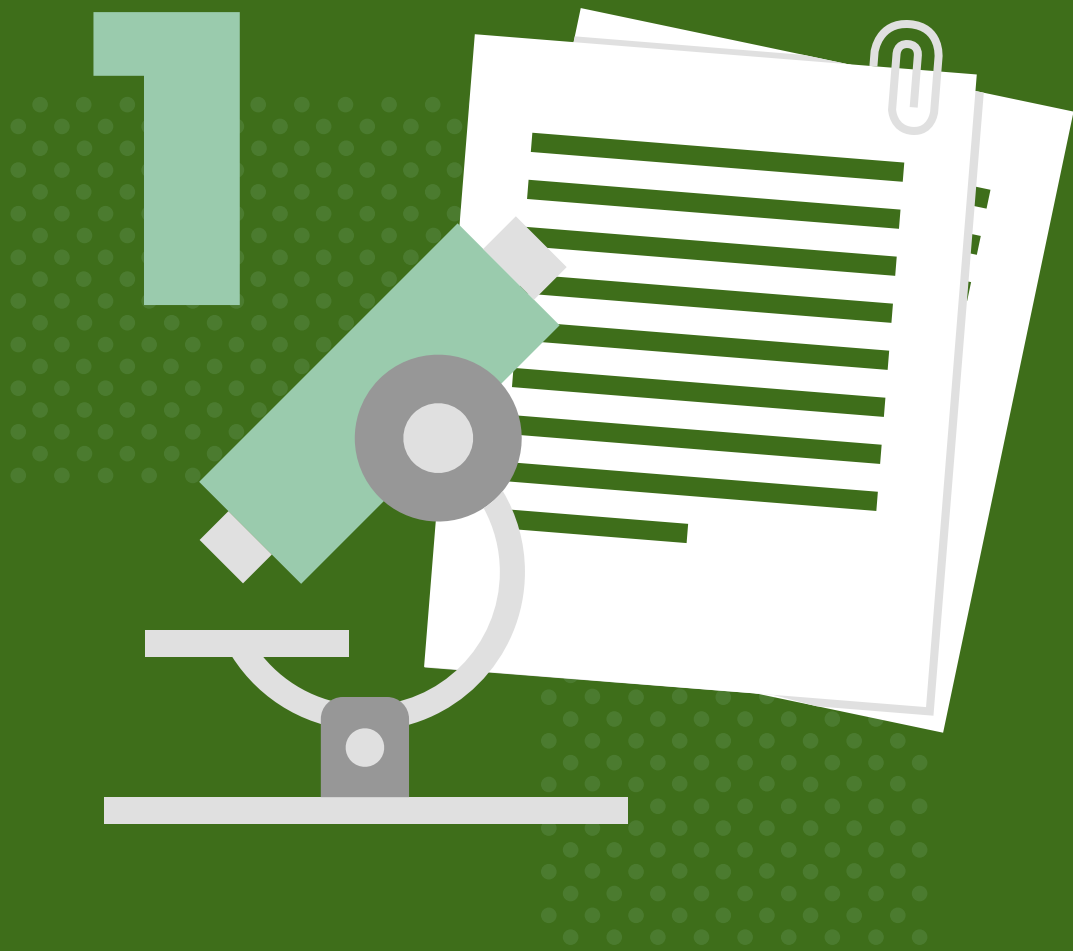
Policy prototyping is a methodology to assess the efficacy of a policy by testing it in a controlled environment first. Policy prototyping applies the design thinking approach, which is common in product and service design, to the development of law and policy.

Legal philosopher Lon Fuller has defined law as the enterprise of subjecting human conduct to the governance of rules.⁴ Rules (laws and policies) are made to influence the behavior of individuals, groups, or organizations (the norm addressees) with the goal of bringing about certain mutual behavior, action, or abstention from action. In other words, a law or a policy is a means to achieve a particular policy goal.

While laws and policies are discussed and consulted, they are seldom tested in practice. As such, laws are typically enacted without it being clear whether they are actually effective and "fit for purpose." We want to test whether the AIA is effective in achieving the goal for which it has been created. To this end, we determine the overall goal(s) of the AIA, examine the required conditions (outcomes) necessary to achieve these goals, and assess whether the policy interventions proposed in the AIA actually contribute to creating the desired conditions.⁵

The idea is that policy prototyping will lead to more effective policymaking, thereby avoiding the societal costs of "bad policy." These costs can be not only of an economic nature (e.g., high compliance costs, high enforcement costs, or loss of opportunity) but also of a non-material nature such as infringement of human rights.





Research goals and methodology

Below, we describe our methodology, starting by setting out the goals for the research and then describing the method used to achieve our research goals. We also discuss the scope and limitations of this research.

This Open Loop program was focused on getting quantitative and qualitative input from companies that will likely be subjected to the requirements of the AIA. Over a period of three weeks (from June 2022 to July 2022), participants were invited to provide feedback and input on selected provisions of the AIA. A combination of multiple-choice and open

questions based on five themes was used to allow participants to express their views, which are described in more detail below. The aim was to gather feedback on key aspects of the AIA and to test whether the proposed provisions in the AIA would contribute to achieving the overall goals of the AIA.

To assess the (potential) effectiveness of the requirements outlined in the AIA in contributing to its overall goals, we created a series of activities for participants, each focusing on key aspects of the AIA:



Activity 1: Diving into the AI ecosystem – taxonomy of actors

Article 3 AIA

In the first activity, we examined the taxonomy of AI actors provided in the AIA. The legislator differentiates between norm addressees (i.e., providers and users), setting clear rules for each actor depending on their role in the AI ecosystem. We hypothesize that the goal of the legislator here is to assign clear responsibilities to each actor, ensuring that each of them takes a degree of responsibility for ensuring trustworthy AI systems. The question is whether the roles that exist in the real-world AI ecosystem match up with the taxonomy provided by the legislator.



Activity 2: Managing risk

Article 9 AIA

In the second activity, we looked at risk management. The AIA proposes strict obligations that high-risk AI systems must comply with before being put on the market. Specifically, the AIA requires adequate risk assessment and mitigation systems, laid out in Article 9. The policy goal here is to limit the risks of AI systems. The goal of this activity was to assess whether the proposed obligations contributed to limiting risk and whether meeting these requirements was considered realistic by the participants.



Activity 3: Setting up data quality requirements

Article 10 AIA

In the third activity, we quizzed participants on the data quality requirements in the AIA. We wanted to understand whether participants thought they could produce and maintain data sets that would meet the data quality criteria set forth by the AIA and if not, why not. We hypothesize that the goal of the legislator is to ensure the accuracy and robustness of AI systems by mandating that the data used to train AI systems meets sufficient quality thresholds to allow for accurate predictions.



Activity 4: Drawing up the technical documentation

Article 11 AIA

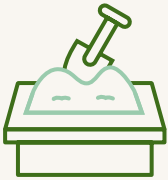
In the fourth activity, we looked at the technical documentation requirement. We hypothesize that the goal of the legislator is to let the provider draw up technical documentation that allows for independent (i.e., third party) assessment of whether the AI system meets the conformity requirements set by the EU. We tested whether participants were able to draw up this technical documentation.



Activity 5: Assuring transparency and human oversight

Articles 13 and 14 AIA

In the fifth activity, we looked at the transparency and human oversight requirements. We hypothesize that the goal of the legislator is to ensure that humans can assess how an AI system is operating by interpreting its outputs. This presupposes that the human responsible for the oversight has a particular skill level. We quizzed the participants on the level of skill required to be able to provide this oversight.



Bonus activity: Regulatory sandboxes

Article 53 AIA

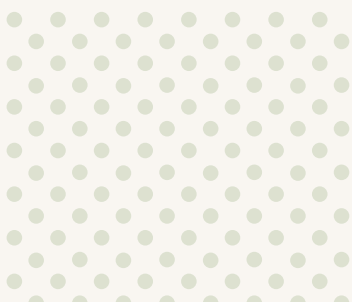
Finally, for those interested, we included a "bonus activity" on regulatory sandboxes.⁶ In this activity, we asked the participants whether they would be open to participating in the proposed regulatory sandboxes of Article 53 AIA and what conditions they felt must be met to foster innovation in regulatory sandboxes.

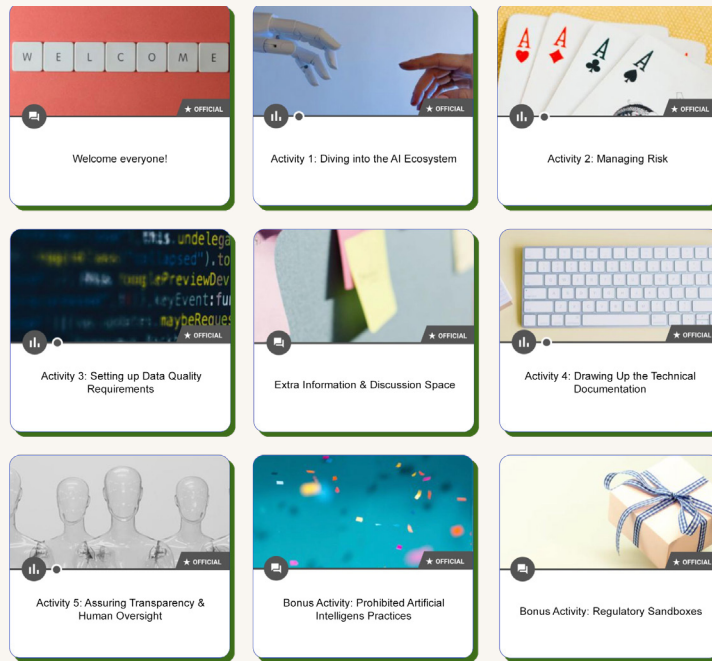
The Open Loop Forum

To gather feedback from participants, we formed a dedicated Open Loop Forum (OLF), which was a closed online platform where invited participants met to discuss topics and take part in various research-related tasks. The activities were moderated by the Open Loop program team, who interacted directly with the participants by asking follow-up questions for clarification and expansion. The platform enabled our team of moderators and researchers to use a mix of qualitative and quantitative methods to uncover insights from participants,

who were recruited according to certain target criteria (see below).

The OLF allowed us to connect a large number of stakeholders affected by the new policy, and to integrate their feedback and ideas early in the process. It helped us learn and improve the program outlines, publications, and recommendations by identifying participants' insights, needs, wishes, and concerns. Within the OLF, we built strong relationships with a user community to establish a trusting, co-creation mindset.





Participants

To perform the task at hand, the research team recruited participating companies and correspondent representatives to give feedback on selected articles of the AIA (hereinafter "participants").⁷ Participants were selected based on two criteria: i) whether they develop AI/ML products that could fall under the scope of the AIA; and ii) whether they are active/operating in the EU market. The sample of selected participants ensured a balanced regional coverage of AI companies (Southern Europe, Central and Eastern Europe, Western Europe, and Northern Europe), but also included companies that are not based in the EU but that operate in the EU internal market. The majority of participants joined the program via personal invitation. Spontaneous applications received via the Open Loop website and social media channels were also accepted. When participating in the program, participants did not disclose any proprietary information and/or sensitive data.


A total of 62 users from 53 companies participated in the program. There were a total of 2966 posts (answers to multiple choice questions and open discussion).

There were 4 large companies (>250 employees), 7 medium-sized companies (<250 employees), 26 small enterprises (<49 employees), and 16 micro enterprises (<10 employees). Most of the companies were startups or scale-ups (having been in business for < 5 years). The companies operated in a variety of what could be considered high-risk and non-high-risk sectors, such as law enforcement, healthcare, education, HR, logistics, finance, and marketing. The companies were active throughout the EU, with most of them being active in Germany, Italy, and Spain.

Most of the participants had a technical background or role (e.g., data scientist and product manager). Given the size of the companies, most of the participants also performed managerial roles (e.g., CEO and CTO). About half of the participants had over 5 years of experience in the field of AI. Participants were evenly split in terms of compliance experience with policy and regulations in the digital field, with about half indicating that they had zero to basic experience and half indicating that they had considerable to advanced experience.



62 users
from 53 companies
participated in the program


























2966 posts
(answers to multiple choice
questions and open discussion)







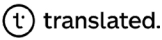










To thank them for their participation, companies were offered the opportunity to attend guest talks and Q&A sessions with Meta’s PyTorch and AI Research teams, and – for selected companies taking part in the second phase of the program – to receive ad credits on Meta’s platforms. Participants are also featured on Open Loop’s website and in the dissemination materials of the program. At the end of this program’s phase, participants gathered for a virtual get-together – an informal meeting where they were given the opportunity to network, provide feedback on the program,

and suggest improvements. Participants highlighted that Open Loop was a learning opportunity. In effect, those who did not have the chance to read the AIA were able to delve into certain aspects of it and consider its potential impact on their business thanks to the program. Participants also appreciated how complex legal information was conveyed with the right amount of detail, allowing them to give meaningful and informed answers. Finally, participants suggested making the OLF more interactive, enhancing ways for participants to learn from each other.

Overview of Participating companies

Company	Country	Industry/Area of Activity
	Denmark	AI governance solutions
	Italy	AI solutions, Cybersecurity
	Spain	Conversational AI
	United States	AI governance solutions
	Poland	AI software development
	Ireland	AI governance solutions
	Canada	AI governance solutions
	France	Conversational AI
	Italy	AI solutions
	Netherlands	AI governance solutions
	Georgia	Cosmetics & Retail Recommendation Systems and Application Tracking Systems
	Italy/United Kingdom	Pricing and supply chain optimization

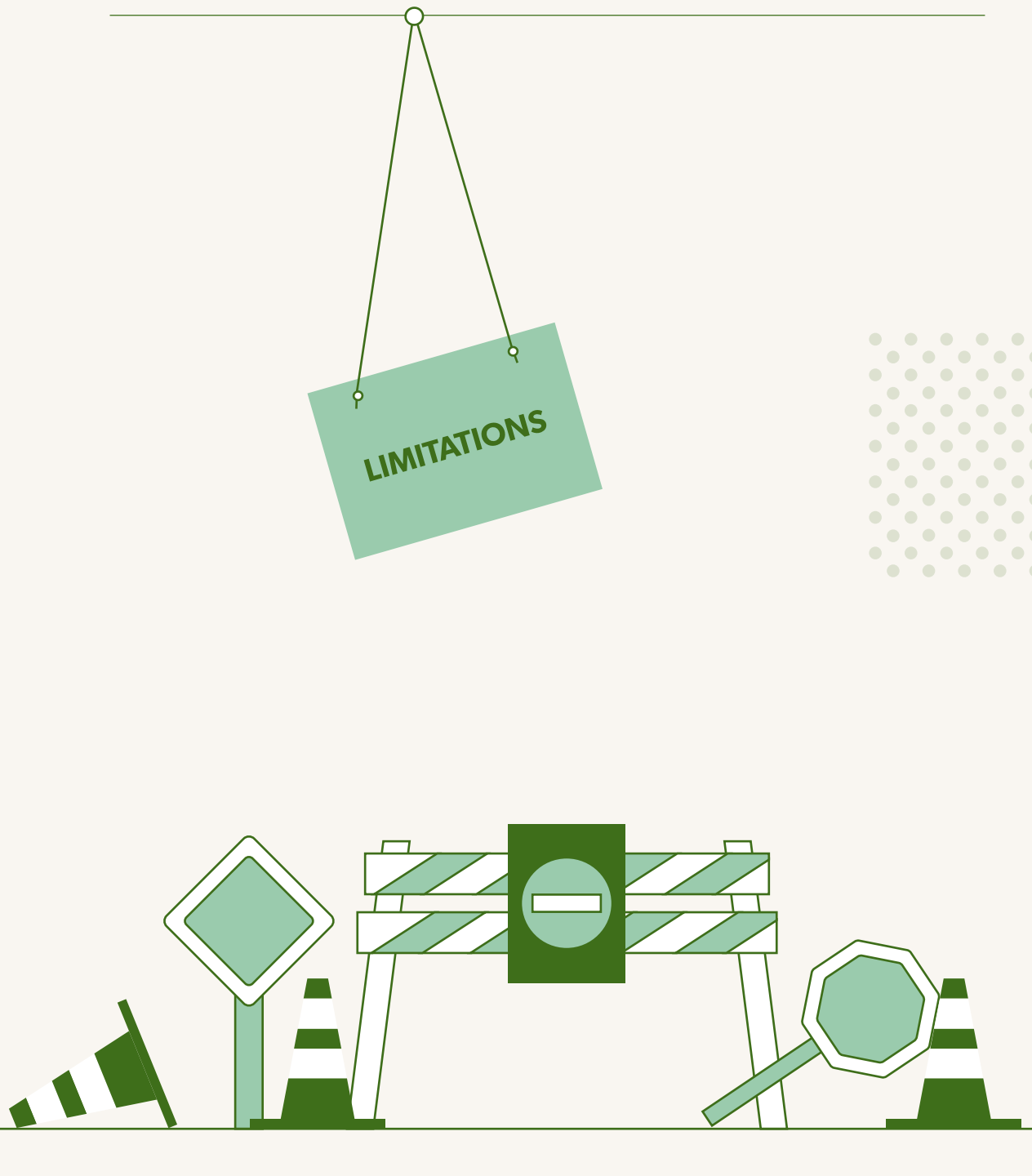
 <p>United Kingdom Media</p>	 <p>Netherlands AI solutions</p>
 <p>Greece Telecommunications</p>	 <p>Switzerland Healthcare</p>
 <p>Italy Conversational AI</p>	 <p>Italy E-commerce, VR</p>
 <p>United Kingdom/Austria Transport, Logistic, Blockchain</p>	 <p>Greece Computer Vision</p>
 <p>Spain Cloud Analytics Solutions</p>	 <p>United Kingdom/Italy/Singapore Cybersecurity, AI solutions</p>
 <p>Germany Responsible AI testing</p>	 <p>United Kingdom Education</p>
 <p>Germany Healthcare</p>	 <p>Italy AI solutions, IT consulting</p>
 <p>Italy AI solutions</p>	 <p>Italy Human Resources</p>
 <p>Germany Robotic solutions</p>	 <p>Germany NLP</p>
 <p>Portugal Media, Events</p>	 <p>Germany Transport, Logistics</p>
 <p>Italy Manufacturing</p>	 <p>Germany Learning</p>
 <p>Germany AI platform</p>	

 <p>Netherlands Banking, Insurance, Cybersecurity, and Critical Infrastructure</p>	 <p>Germany AI-powered Emotion Analytics and Eyetracking</p>
 <p>Lithuania AI solutions</p>	 <p>Lithuania/Estonia/Latvia Legal services</p>
 <p>Portugal Media</p>	 <p>United Kingdom VR, Education</p>
 <p>Italy Language services</p>	 <p>United Kingdom Legal Services</p>
<p>VEDRAI</p> <p>Italy AI solutions</p>	 <p>Portugal/United States/ United Kingdom/Spain VR, Healthcare</p>
<p>VISOR AI</p> <p>Portugal Conversational AI</p>	 <p>Portugal Telecommunications</p>
 <p>France Speech AI</p>	 <p>Lithuania Education</p>
 <p>Netherlands Conversational AI</p>	 <p>United Kingdom Transport, Logistics</p>
 <p>Portugal/United States Transport, Logistics</p>	 <p>Portugal AI Solutions</p>
	 <p>United Kingdom/India AI-risk transparency platform</p>

Scope and limitations

The scope of this research is limited to specific provisions of the AIA. In particular, we explored Article 3 (definitions), Article 9 (risk management), Article 10 (data and data governance), Article 11 (technical documentation) and the associated annexes, and Articles 13 (transparency) and 14 (human oversight) of the AIA.

By exclusively involving AI companies as participants, this research may be susceptible to limitations. Given that these companies will likely be subjected to the AIA, there may be a certain bias in the answers given.⁸ Furthermore, participants may over- or underestimate their ability to comply with particular requirements of the AIA in their answers. Finally, given the sample composition – most of the participants may be considered either startups or scale-ups – findings may not be representative of medium and large companies.





Activity 1

Taxonomy of AI actors

In this activity, we focused on the AI ecosystem and taxonomy of AI actors. The AIA distinguishes between different types of norm addressees (mainly between providers and users). It is interesting to see whether these accurately reflect the actual actors in the AI ecosystem.


To test whether the taxonomy of AI actors in the AIA is accurate, we asked participants to:

- 1 identify their own role under the AIA and reflect on the completeness of the taxonomy of AI Actors,
- 2 reflect on their own role in the AI ecosystem and the distinction between different actors, and
- 3 do an exercise to determine whether they could correctly assign the definitions of the AIA to different actors in a scenario.

Furthermore, we asked the participants whether they could accurately describe the purposes of their AI system.

Task 1:
Determining your role under the AIA

For this question the participants were tasked with reading the definitions of the different actors outlined in the AIA:



Article 3

Definitions

1

'provider' means a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge;

2

'user' means any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity;

3

'importer' means any natural or legal person established in the Union that places on the market or puts into service an AI system that bears the name or trademark of a natural or legal person established outside the Union;

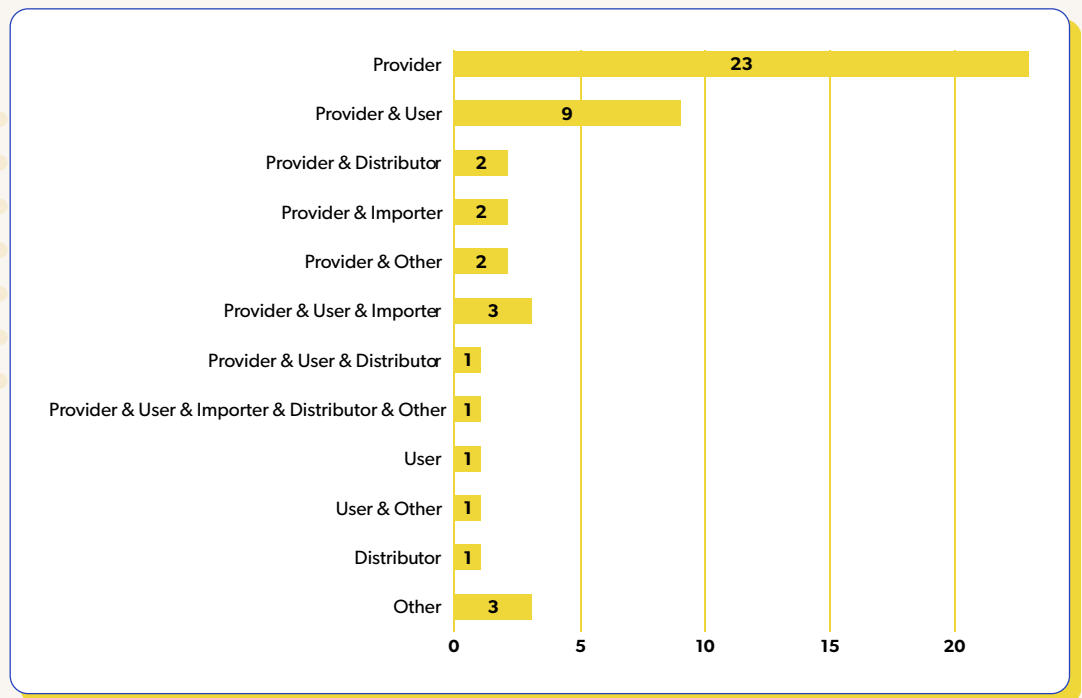
4

'distributor' means any natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market without affecting its properties;

Article 3(1-4) AIA

They were subsequently asked which role(s) applied to them.

Respondents were able to select more than one answer. This yielded the following distribution:



According to the definitions given in the AI Act, which of the following actors are you? [N = 49]

In general, it seems clear for respondents which role or roles they have.⁹ Most respondents indicated "provider" as their role, some also "user." This implies that the AIA is clear in its definition of what a user and a provider is.

What is interesting to note is that most respondents that indicated "user" as their role also indicated "provider." Only one respondent indicated "user" as their only role.¹⁰ This suggests that providers of AI systems are often also users of AI systems because they themselves need AI components to operate their AI systems. Indeed, some of the providers indicated that they also use other AI systems as part of their own system:

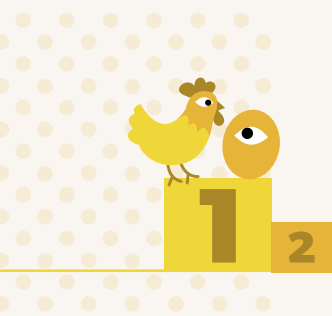
"We also use different AI-based systems ourselves, especially during the research and development of our own products."

TAWNY

This results in a sort of AI "chicken and egg" situation where it is difficult to neatly disentangle provider–user relationships, as most of the time these actors will assume both roles. This is relevant from the perspective of the AIA as these dual roles may create risks downstream insofar as a provider might rely on components/AI systems provided by other providers. When one of these components/AI systems is faulty, this might negatively impact the user, individuals, and society at large. While providers of high-risk AI systems must meet strict requirements to ensure the correct operation of their AI system, this does not seem to be the case for (non-high-risk) "upstream providers." For instance, when an upstream AI provider provides a machine learning model to an AI company, this will not be considered "high risk". How-

"(...) We rely on AI models developed by others, in which capacity we seem to be a 'user'."

LearnerShape



ever, when the AI company builds an AI system using this machine learning model and sells it to users in law enforcement, this system will be considered high risk. This raises the question wheth-

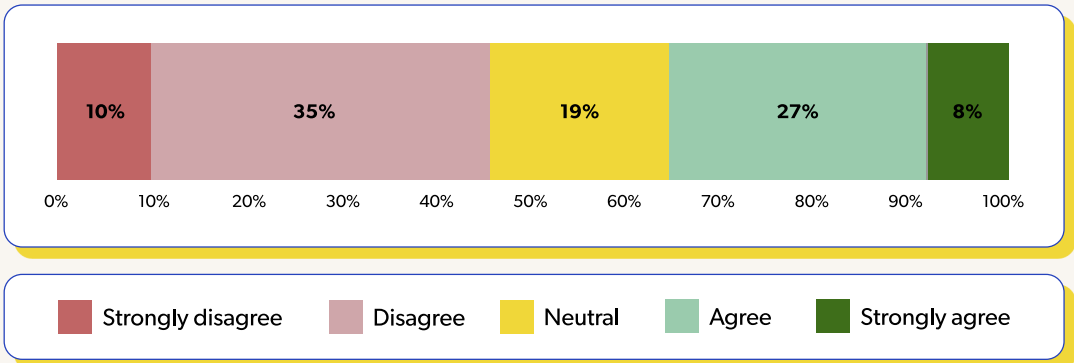
er the upstream provider should meet the same requirements as the final provider or if it is upon the user of the component/AI system to assess whether it is safe to use in the high-risk AI system.

Task 2:
User and provider

The participants were asked to respond to the following statement:

In the AI ecosystem, there is always a clear distinction between a user and a provider.

The distribution of the responses was as follows:



What is your opinion regarding the following statement? "In the AI ecosystem, there is always a clear distinction between a user and a provider." [N = 48]

When asked whether the distinction between a user and a provider is always clear, 45% of the respondents disagreed. It seems that a vast number of AI systems make use of other components and services, which makes the participants "provider" as well as "user." The answer to this question seems to support the idea that while the definitions of "provider" and "user" are clear, at least "on paper," in reality there is not always a very clear distinction between the two. First of all, a provider may also be a user. Second, and we will explore this further below, there may be cooperation between the provider and the user to such an extent (e.g., the user

sets requirements, provides training data) that it becomes less straightforward to determine who should ultimately be responsible for the final product:

"The data feeds into the system and the system delivers an output. From this output, other data is produced, collected, and fed into the broader data set. User and provider can become intertwined."

Qubit Ventures

Task 3:
Intended purpose of the system

In the third exercise, we asked the participants to look at the definition of "intended purpose" from Article 3(12) of the AIA:



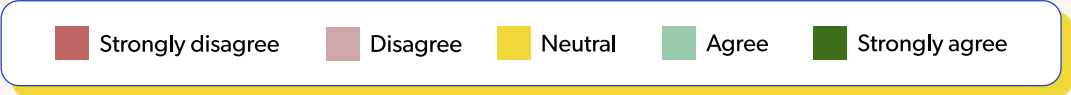
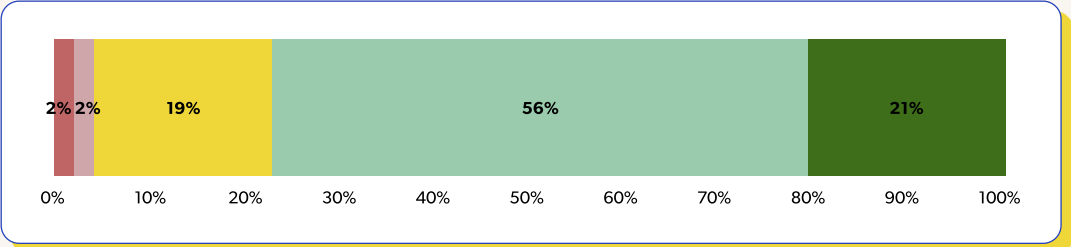
"Intended purpose" means the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation.

Article 3(12) AIA

We then asked them to respond to the following statement:

Based on the definition, I would be able to describe the "intended purpose" of the AI system we provide.

Notably, 77% of the respondents indicated that they agree or strongly agree with the statement. As such, there is little doubt that providers will be able to meet this requirement under the AIA.



Based on the definition, I would be able to describe the "intended purpose" of the AI system we provide. [N = 48]

Task 4:
Scenarios
on actors

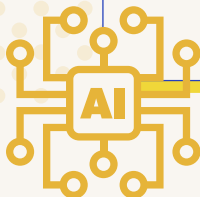
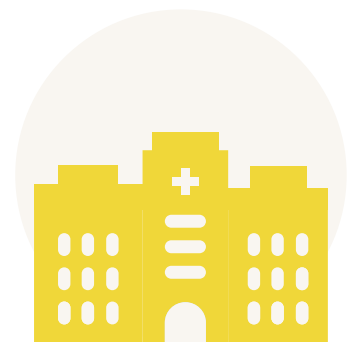
Finally, the participants were provided with two scenarios. In the first one, they had to assign a role to each actor described in the scenario (i.e., which role in the taxonomy of AI actors from the AIA applied to them).

Scenario 1:

A machine learning model is placed on the market by the company ABC. This model is used by a hospital.

Based on the taxonomy provided in article 3 of the AIA, what would be the roles of ABC and the hospital (provider vs. user)?

Please explain briefly why.



The most common opinion was that **the hospital is the user and ABC the provider.**

"I'm going to assume (in the absence of any further context) that ABC developed the system, placed it on the market under its own name, and is not using the system itself, therefore fitting the definition of 'provider' under the act."

Enzai Technologies Limited

However, participants noted that there are many factors that can influence the answer. In particular, if the user (the hospital) makes changes to the (intended purpose) of the system or is involved in determining training and input data:

"ABC provider, Hospital user, however the lines can easily become blurred if the hospital starts inputting data that is clearly not within the intended use by ABC for its model, or it starts to embed the ABC AI into an in-house developed AI pipeline."

SynerScope bv

"Agree that ABC is the provider and the Hospital is the user, assuming the hospital won't be feeding any new data for the model training/update."

YData Labs Inc

"I'm also going to assume (again, in the absence of context) that the hospital had no involvement in the development process, acquired this off-the-shelf from ABC, and is using the system only for itself (i.e., not providing to patients), therefore fitting the 'user' definition under the act."

Enzai Technologies Limited

Participants also noted that the binary provider–user taxonomy belies the complexity of providing and using AI systems in practice:

"From a legal perspective, ABC would be in that scenario the 'provider' for the hospital (the 'user'). From a personal level, there would be multiple user–provider relationships along the value chain. E.g., the data scientist at ABC is the provider of a model which is used by the doctor working in the hospital, but from a patient perspective, the hospital is the provider incl. the responsibilities which come with it."

Deeploy

The concept of the "end user" of an AI system (i.e., an operator) seems to be missing from the taxonomy of AI actors. For instance, a hospital (the user) may buy an AI system, but that system will be operated by, e.g., a doctor or a nurse (the operator). Furthermore, the person who is subjected to a decision/prediction of an AI system (e.g., the patient) also does not seem to be identified in the AIA.

Finally, concerns about the allocation of responsibilities along the value chain were raised:

"So perhaps more clear where the liability sits here (with ABC) could well depend on where ABC got their training data and whether that included inherited liability for error."

Travers Smith

The issue of downstream responsibility as mentioned above, or the issue of shared responsibility between the provider and the user, does not seem to be explicitly covered in the AIA. The binary approach, which seems to be borrowed from product liability – where there is a more clear distinction between the user and the provider – might lead to questions about



its applicability in real-world scenarios. While these might be solved in practice, for instance, through litigation, it would be beneficial for the addressees of the AIA, and for all different types of entities involved in the AI value chain, if the legislator provided a clearer position on this point. This could be done, for instance, by defining under which circumstances a user and a provider cooperating in building an AI system become "co-providers." The need for legal clarity around the taxonomy of AI actors

in the AI value chain becomes even more urgent in light of the two recent proposals on liability put forward by the European Commission (namely the AI Liability Directive and the proposal for a Revision of the Product Liability Directive).¹¹ While the former proposes harmonized civil liability rules for individuals harmed by AI systems and adopts the same definitions of the AIA, the latter equates providers to manufacturers but also introduces the concept of "economic operator."

After this scenario, the participants were confronted with a more complex scenario:

Scenario 2:



A multinational e-commerce company (Aladin) hires a consulting firm (Baloo) to build a machine learning model for selecting the best candidates for job offers. The consulting firm (Baloo) uses an open-source machine learning framework (Pocahontas) provided by a social technology company (Goofy) for training the model. The training data is provided by a multinational data-based marketing company (Simba). The model turns out to be biased against women, most likely because the training data was not representative.

In this second scenario, the participants were asked to determine who was liable for the biased AI system. Below is an overview of the opinion of participants on the responsibility of each of the parties involved:

Aladin (user)

Aladin was mentioned the most when it came to liability as they commissioned and used the system.¹² As the user, they have an obligation not to harm those subjected to the machine learning model (the job applicants):

"The user (Aladin) is always liable for any damages caused to the end users; they should respect the terms of the contractual (ethical, good faith, etc.) agreement regarding the services or products provided from the user to the end user."

Irida Labs

However, some respondents argued that Aladin should not be held liable. For example, because they may lack the internal knowledge regarding the AI system:

"[Aladin] couldn't be responsible for everything related to the technology they contract, they couldn't possibly know every question on models, training, or data sets."

AUNOA

Baloo (provider)

The second company that was indicated as potentially liable was Baloo. Respondents remarked:

- Baloo is liable if they have received correct specifications from Aladin but provided a biased model.
- Baloo is liable as they provided a biased model (if the bias was identifiable from a data profiling/validation point of view).
- Baloo is liable as they should check the quality of the training data and test for any bias in the prediction from its trained model.

The participants argued that Baloo must always check whether the model is fit for purpose before they release it to the client. The fact that they were provided with biased data could be a mitigating circumstance, but at the same time Baloo should have identified this risk at the data selection/preparation stage.

Simba (data provider)

Some argued that Simba is also liable because they provided a biased data set:

"My special attention goes to Simba as they work data to training-data (for a profit) and they could introduce bias through their work big time."

SynerScope bv

Goofy, social technology company providing Pocahontas

While not many argued that Goofy should be held responsible for providing the open-source framework Pocahontas, questions were raised as to the responsibility for any flaws in open-source products:

"What if Pocahontas is aware of an inherent flaw in its open-source framework, that it hasn't alerted any users to, which it knows can result in dangerous predictions? I appreciate it's OSS and provided on an as-is basis,¹³ but it seems unfair that it hasn't alerted users to this issue."

Enzai Technologies Limited

The participants also mentioned shared liability and argue that there should be a fair distribution of liability:

"Aladin and the vendors in the supply chain Baloo and Simba are liable for the biased results of the machine learning model. These risks could have been mitigated if Simba had conducted a data quality assessment and had included more representative data on women applicants. Before deploying the model, Baloo should have run a fairness assessment to check for disparate impact. Aladin ideally should have had an independent third party conduct a periodic bias audit to publish findings. We recommend that Goofy should outline system cards and bias metrics on Pocahontas to help the open-source developer community understand the inherent limitations of the framework."

Zupervise

Furthermore, they questioned whether the AIA should regulate liability at all:

"I'm not 100% sure about the details of how liability can traverse the supply chain in other, more established fields (e.g., car manufacturing), but I don't see why the AI field should be different. I think it mostly depends on the contracts between the entities."

TAWNY

"Should the regulator overrule a b2b contract? As long as liability sits somewhere."

Travers Smith

Some participants pointed out that certification mechanisms, standardization, and adherence to codes of conduct could not only help limit risks of AI, but also create clarity on who must be held liable.

"Title IX of the EU AI Act creates a framework for the creation of codes of conduct, which aims to encourage providers of non-high-risk AI systems to apply the requirements of the EU AIA voluntarily. Aladin and all other players should enforce the use of those codes of conduct."

Palko

When it comes to the attribution of liability, civil and commercial law already provide a system for its allocation. Companies can assign and off-load risk via contractual arrangements. Resorting to contracts carries the benefits of being more flexible and tailored to the specific circumstances of the case.

In the case of litigation between parties, it is likely that the judge will assess whether the parties have acted according to what objectively may be expected of them. Here, it might be helpful if the AIA would be more clearer on what the responsibilities are when parties cooperate (provider and user) in creating an AI system, and/or what the responsibilities are of parties providing components or data to providers of high-risk AI systems.

Standardization and certification bodies, by developing technical standards for the certification of AI systems, may prove to be helpful in determining liability. Standards and certifications may contribute toward creating an objective standard of what is considered appropriate conduct for a market player. Through a draft standardization request, the European Commission entrusted the European Committee for Standardisation (CEN), the European Committee for Electrotechnical Standardisation (CENELEC), and the European Telecommunications Standard Institute (ETSI) with operationalizing the technical requirements of the proposal by the time of its entry into force. Based on the responses of our participants, we underscore the importance of these two instruments working in tandem.





With regard to the taxonomy of AI actors from the AIA, we can establish that the definitions of provider and user are sufficiently clear for the participants, at least "on paper." There is some confusion about the term "user", as participants also associate that with the person who is the subject of an automated decision (e.g., the patient in the first scenario). In the taxonomy of AI actors in the AIA, this "subject of an AI system" is noticeably absent.

While the definitions of user and provider are clear, most participants argue that in practice their roles are not mutually exclusive. A provider is often also a user (as it may use other AI systems for the functioning of its AI system). Furthermore, the user may significantly influence the work of the provider, for instance, by setting requirements for the AI system and/or providing training data. This means that, in practice, it is much more difficult to assign roles to each of the parties involved in the creation and use of an AI system.

This in turn affects discussions on the attribution of responsibility along the value chain. In the above-mentioned scenarios, we observe that the interdependence of different actors in the AI ecosystem raises questions regarding who should be held responsible, as actors are dependent on one another for, e.g., training and input data, models, and other technologies. What we may tentatively conclude is that the binary approach of the AIA does not fully match the reality of the AI ecosystem. Unlike traditional products such as forklifts or toasters, the user of the AI system is often also involved in the creation of the AI systems. Users, for instance, commonly provide relevant data sets for training while also setting requirements for the performance of the system. Furthermore, while this is not unique to AI, AI systems may depend on components of third parties or interact with the data of third parties.

The limitations of the AIA proposal have been partially addressed by the European Parliament, with several amendments calling for a more granular taxonomy that would allow for a more appropriate allocation of responsibilities. For instance, the compromise amendments presented by the JURI committee in its opinion attempt to clarify the responsibilities of providers as well as users, specifying under what circumstances responsibilities might shift to another actor.¹⁴





Activity 2

Managing risk

The AIA proposes strict obligations that high-risk AI systems must comply with before being put on the market. Specifically, the AIA requires adequate risk assessment and mitigation systems, laid out in Article 9. This article demands a continuous iterative process that runs throughout the entire lifecycle of a high-risk AI system (targeting risk identification, evaluation of risks, adoption, and testing of risk management measures).

This activity aims to gain insight into whether providers of AI systems are willing to do a risk assessment when they are not classified as high risk, to understand if AI companies can meet all the criteria mentioned in Article 9(2 AIA), and whether participants think they can assess and predict risks and misuse.

For this activity, it is relevant to note that about 75% of the participants had basic, considerable, or advanced experience with complying with regulation in a digital context. Of the companies involved in this Open Loop program, most either did not have a dedicated risk management system for AI or were in the process of developing one. However, some companies indicated that they had specific risk management systems for privacy and data protection (GDPR compliance).



Article 9 Risk management system

1 A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems.

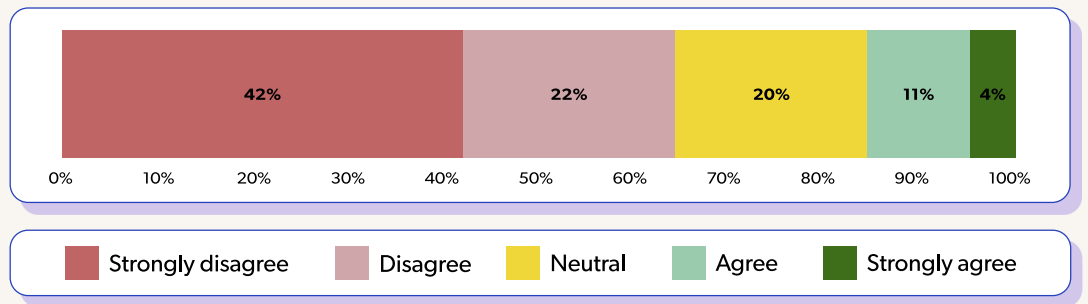
2 The risk management system shall consist of a continuous iterative process run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic updating. It shall comprise the following steps:

- (a) identification and analysis of the known and foreseeable risks associated with each high-risk AI system;
- (b) estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse;
- (c) evaluation of other possibly arising risks based on the analysis of data gathered from the post-market monitoring system referred to in Article 61;
- (d) adoption of suitable risk management measures in accordance with the provisions of the following paragraphs.

Article 9(1-2) AIA

Task 1:
Risk assessment
appetite

Participants were asked if they agreed that they would not perform a risk assessment if their systems were not classified as high risk according to the AIA. The results show that this is not the case for most participants. In fact, a majority of the respondents disagree that they would not perform a risk assessment.



If my system is not classified as a high-risk system according to the AI Act, I won't perform a risk assessment. [N = 45]

When asked to explain their position, participants noted that non-high-risk systems can still entail some risk or interact with high-risk systems in the value chain. Others noted that performing a risk assessment might also increase trust in AI with clients:



"It would also give potential clients peace of mind knowing correct safety protocols are being followed. Everything has a risk and anything which is performing automated tasks, no matter how trivial they may seem, should have some type of risk assessment conducted."

The VR Hive

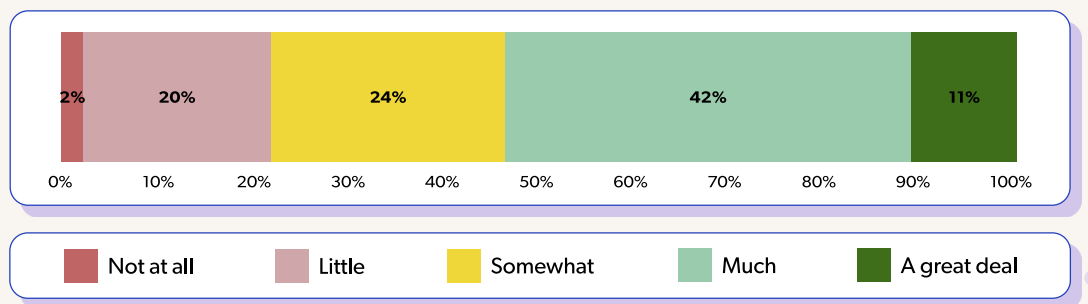
There are also participants that stated that they will not perform a risk assessment if they are not compelled to by law. That means that they will not perform a risk assessment when they are not classified as high-risk AI. We have seen several arguments for not performing a risk assessment for non-high-risk AI systems:

- If an AI system does not pose risk to natural persons, a risk assessment should not be mandatory.
- A risk assessment should be mandatory for all, or it should not be mandatory at all.

Finally, one of the participants doubted that when voluntary risk assessments are conducted by companies, they would happen in a formalized, structured (and resource-intensive) way.

Task 2:
Understanding
known and
foreseeable risk

Participants were then presented with risk management requirements from the AIA. To assess whether the proposed risk management requirement is feasible and clear, we asked participants whether they understand what is meant by "known and foreseeable risks" as mentioned in Article 9(2)(a). The responses show that the majority of the participants understand what is meant by the phrase "known and foreseeable risk." We asked these respondents to provide examples of these risks.



Do you understand what is meant by "known and foreseeable risks"? [N = 45]

The responses fall broadly into these categories:

- biased data in general;
- gender, discrimination and racial bias;
- wrong decisions;
- inability to detect unwanted behavior of the model;
- using an AI system that has either not been tested before at all or in a different environment; and
- performance issues.

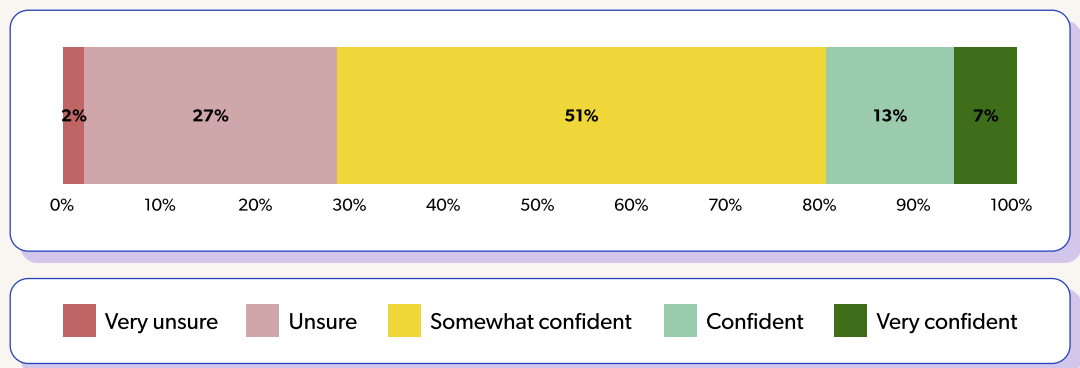
A substantial number of the respondents have little or no understanding at all about what is meant by the phrase "known and foreseeable" risk. We asked them to elaborate on what, in their opinion, is unclear. The responses vary widely. Some wonder whether an overview of (categories of) risk exist that could help in this exercise. Others feel that risks are subjective, especially when it comes to determining what should be known and foreseeable. One participant notes:

"'Known and foreseeable' to whom? To a regulator? To a PhD-level data scientist? To a responsible AI expert with a background in sociology? We would advocate for much greater specificity around exactly what risks are expected to be assessed, because each person comes to the table with a different level of understanding related to AI risk. It's critical to be clear about the dimensions of risk that are important to measure and manage."

Credo AI

Task 3:
Confidence in assessing risks and misuse of AI systems

We then asked participants whether they were confident that they were able to comply with the second element of Article 9(2) "assess and foresee risks and misuse of your AI system." A majority is somewhat confident.



How confident are you that you can assess and foresee risks and misuse of your AI system? [N = 45]

Respondents shared a wide range of concerns. Some struggled with the phrase **"foresee,"** as some risks can be obvious in retrospect but very hard to identify before they materialize. One participant noted that misuse is hard to predict for all users and proposes limiting risk assessment to that what is seen as objectively reasonable:

"However, there will always be certain risks that are foreseeable and should be caught during development. The 'reasonable man' test under English law, despite its inherent flaws, is probably the best test we have to analyze this."

Enzai Technologies Limited

Many noted that it is difficult to foresee expected uses of their AI systems when they are used or integrated in third-party applications:

Our alpha and beta testing has shown (as is well-known) that user behavior is difficult to predict. Also, because we build infrastructure for third-party applications, it is (by design) difficult to foresee many of the expected uses of our infrastructure."

LearnerShape

One participant remarked that envisioning misuse is an exercise that requires a lot of creative thinking and judgment that is very hard to capture systematically. This thought led a

different respondent to identify the need for a proactive approach:

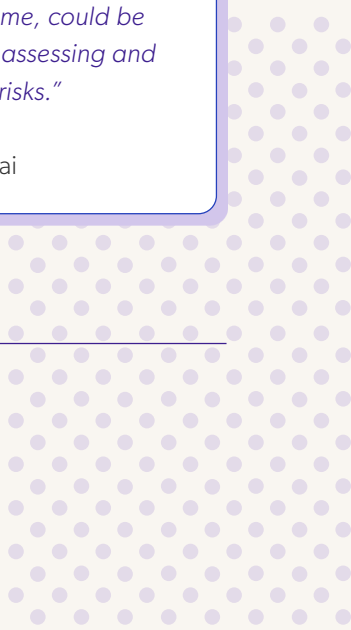
"This is similar to the GDPR policy that limits what data can be collected, how long it can be stored, and what its intended purpose is. The EU is pretty vague about these uses and we will probably need to see some challenges and court cases before it is made clear. I think a proactive approach would involve regular assessments and having documentation continually updated and published to the public."

Vixtape

Respondents that shared that they were (very) confident about assessing and foreseeing risks were asked how they would approach this exercise. The responses (of this small group of respondents) vary widely. Some simply outsource the problem to an "independent assessment of the model's performance", or feel that with some critical thinking they would be able to make an adequate assessment of the risks:

"Thinking about the different use cases in which my technology could be put in, always planning for the worst outcome, could be a good exercise for assessing and foreseeing risks."

indigo.ai

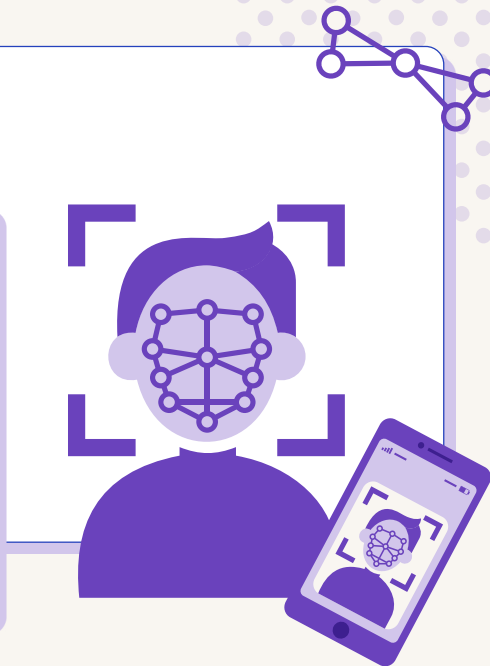


Task 4:
Scenarios on risks

We provided the participants with the following scenario:

Scenario 1:

"Jupiter" AI's algorithm matches faces to a database of a vast amount of images indexed from the Internet, including social media applications. Originally, the intended purpose was solely for law enforcement (i.e., detecting criminals). However, later it turned out that investors and clients used the app for private purposes (on dates and parties etc.).



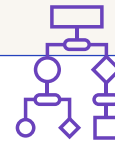
We asked participants to elaborate on whether Jupiter could have foreseen or predicted this risk. The common opinion among the participants was that the risk could have been foreseen because data sets that have been created for one purpose can be used for another. These participants state that the risk of alternative uses of a product should be part of a risk assessment. Specific control or mitigation actions could be designed, e.g., to constrain the use of the algorithm from actors that are not considered law enforcement agencies.

However, participants remarked that, in general, performing a risk assessment is not a trivial task. It takes creativity to assess the risks of alternative uses of a given AI system and even more force of imagination to predict how harmful it can be. Participants argued that while they have the ability to assess risk (and demonstrate creativity in doing so), it is hard to predict all risks, especially when it is unclear which risks are perceived as most dangerous/impactful by the legislator:

"Misuse means outside of intended purpose. This entails a high variety of creative scenarios, which I would not know how to systematically capture. The best I can do is to 1) exercise creative judgment, 2) investigate and collect information on observed misuse, and 3) consult with other professionals, some from the competition, including by monitoring the professional, technical, and scientific literature. This exercise leaves a high chance of focusing on the wrong (less harmful, less frequent) misuses."

Translated

The AIA [Article 9(4)] also mentions that providers must communicate the residual risks that are "judged acceptable." These so-called "residual risks" mentioned in the AIA are risks that remain, even if the AI system is used in accordance with its intended purpose or under conditions of "reasonably foreseeable misuse." To get responses from participants on this particular topic, we provided them with the following scenario:

Scenario 2:

An AI development company (Pluto) is developing a machine learning system designed to predict the risk of criminal recidivism. The system is intended for use in criminal courts throughout Europe, to reduce the time and cost of the administration of justice, and to produce better sentencing outcomes. This sentencing tool is a "high-risk AI system" for the purposes of the AIA.

Company executives know all too well that predictive sentencing systems pose a risk of discrimination on the ground of race and other protected characteristics.

We asked the participants how they would approach assessing the residual risk in this scenario, and what risk should be judged acceptable. Mentioned most often is the rigorous testing of the AI system in different contexts, using a variety of data sets and setting a performance rating threshold. Another useful strategy to identify risk is to assess potential bias in the data sets, subsequently flagging that this is a risk and/or ensuring that the training data is consistent and bias free. Other approaches that have been mentioned are to not rely upon the system for a final decision (i.e., allow for human intervention), or to train the system for a number of years and then compare the decision of the model against new court cases. By measuring the performance of the model, developers would then have a possible idea of the risk. Finally, others suggested assessing the inherent risks of the system minus the mitigation possibilities.

We then asked what residual risks the participants would judge acceptable for this scenario.

Participants focused on the issue of bias in their answers and listed a series of conditions for when a small bias in the AI system would be acceptable:

- if an AI system outperforms human benchmarks,
- if there is a way to appeal the decision or assessment of the AI system,
- if there is a way for someone wrongly sentenced to receive restitution, and
- if the system warns if there is an error so that a human can intervene.

Others only judge residual risks acceptable when it is not used for sentencing itself, i.e., when the system is used to prioritize leads, or if it serves as a workload optimization. Some participants state that they cannot imagine any acceptable risk if a sentencing decision is done without human involvement.



Observations

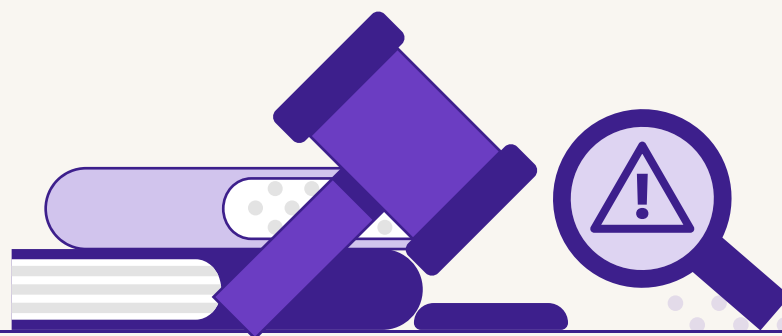
Even though the AIA only mandates formal risk assessment and management procedures for high-risk AI systems, it seems that participants would still carry out risk assessment for their non-high-risk AI systems.

Most of the participants understand what "known and foreseeable" risks are and can identify examples (such as biased data). However, they seem to focus more on the causes of risk (e.g., biased data, inaccurate classification, model degradation, and drift) than on the particular impact of their application on individuals, groups, or society.

It seems that the goal of the legislator is to avoid a negative impact on society. While identifying the cause of a negative impact (e.g., biased data) is important in resolving/avoiding its impact, it might be that risks are underestimated if they are not clearly linked to an associated impact (e.g., what effect does discrimination stemming from biased data have on the individual?). Given that risk is a function of chance and impact, more focus could be given to the impact, rather than focusing mainly on what the root cause of the risk is. Based on this observation, we theorize that those subjected to the AIA (providers and users) might benefit from more concrete guidance on risk and risk assessment in the AIA itself, or through subordinate legislation and/or soft law instruments. While medium or large companies might have the in-house capability for risk assessment, startups and smaller sized companies are unlikely to have comparable capabilities.¹⁵ They would, therefore, benefit from more guidance which could in turn produce better risk assessment and risk management.

An important observation is that it is difficult for participants to anticipate how users or third parties will use their AI systems. In this respect, the requirement to provide a clear description of the intended purpose of the system could be useful. In fact, it can act as a reference point if an eventual harm occurs, as it will be easier to establish if the harmful outcome could have been reasonably foreseen by the provider. While a provider should still bear some responsibility for assessing potential risks that could arise from misuse/abuse by users and third parties, foreseeing all the potential risks posed by misuse/abuse of their systems by users and third parties seems a bridge too far.

More guidance by the legislator (or the regulator) might help in clarifying what "known and foreseeable" risk is and what level of diligence might be expected from those subject to this requirement (mainly providers of AI systems).





Activity 3
**Setting up data
quality requirements**

In this activity, we tested the data quality requirements set in the AIA. The AIA requires providers to use data that

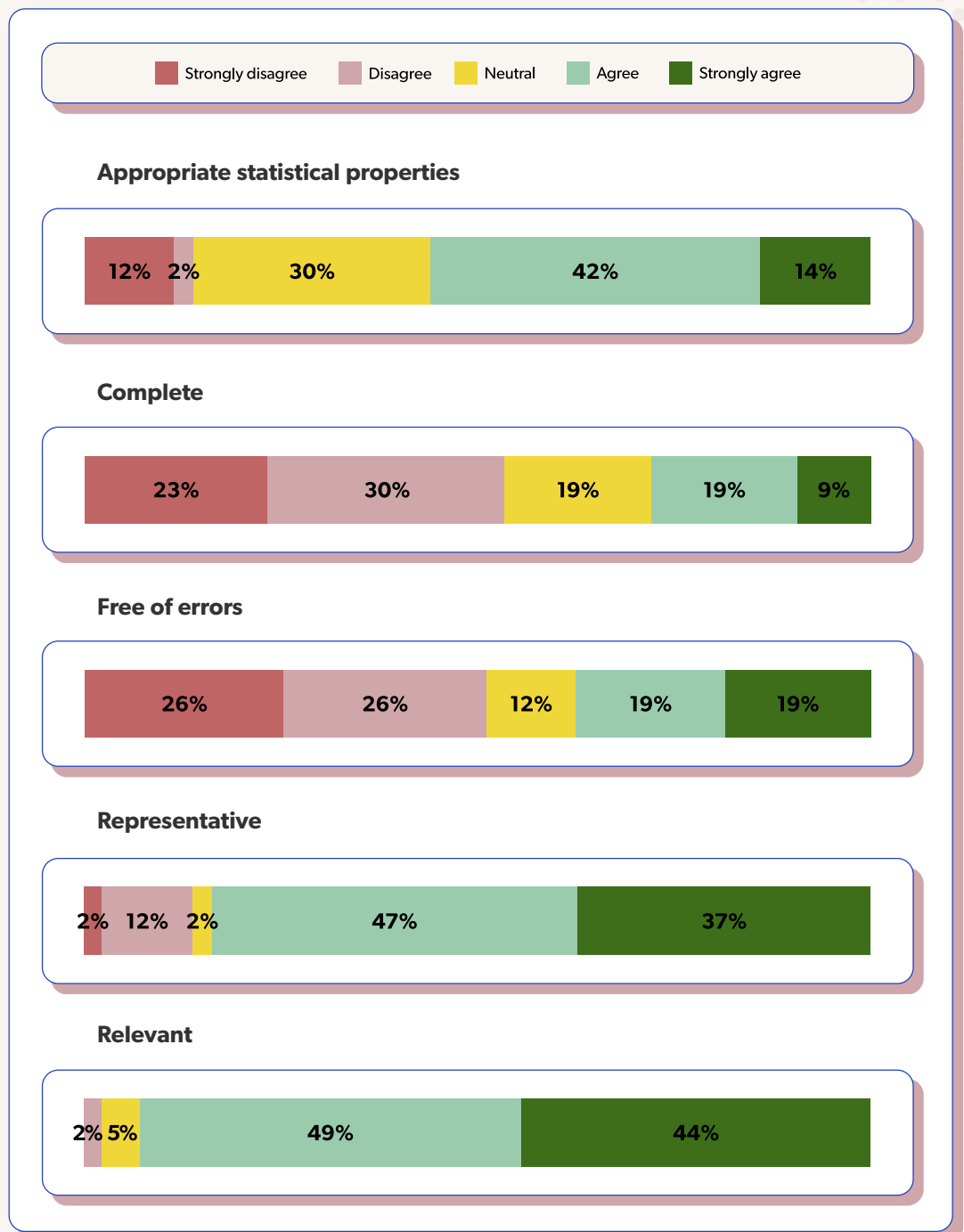
- has the appropriate statistical properties,
- is complete,
- is free of errors,

- is representative, and
- is relevant.

In particular, we assessed whether participants can compile data sets that meet the requirements of the AIA.

Task 1:
Feasibility of data requirements

We asked the participants to share their insights into the feasibility of requirements for data sets used in training, testing, and validation of high-risk AI systems. Participants rated whether they agreed if the following requirements on data sets are feasible:



I think it is feasible to expect data sets that are/have. [N = 53]

The responses clearly show that participants consider some requirements to be less feasible than others. The representativeness and relevancy of data sets are clearly the most feasible. There is less consensus as to whether it is feasible for a data set to have appropriate statistical properties.

The requirements for "error free" and "complete" data sets are clearly perceived as unfeasible. As noted by several participants:

"Unrealistic, and if implemented, seem likely to significantly slow down the development of AI in Europe."

LearnerShape

"Horrible, nothing's perfect, nor are data sets."

Watermelon

"The reality of data means that you will never find a data set that is error-free and complete."

Evo

The requirement of completeness was largely seen by participants as difficult to operationalize. This lack of operationalization means that participants are unsure when completeness would actually be achieved in the AI product development process. Moreover, even if this requirement for completeness were fulfilled, it would only apply to a very specific context, and the entire process to ensure such completeness would need to be redone whenever the system were applied in a new context.

"The specifics of representativeness and completeness require a deep understanding of the contextual use of the intended AI system. Even if a system was trained on a 'complete, representative' data set, if it was used in a new context, then it would no longer be fit for purpose."

Wayve

Participants questioned whether these requirements can be operationalized in an effective manner and noted that these might raise barriers for small companies to enter the market.

Task 2: European Parliament amendments

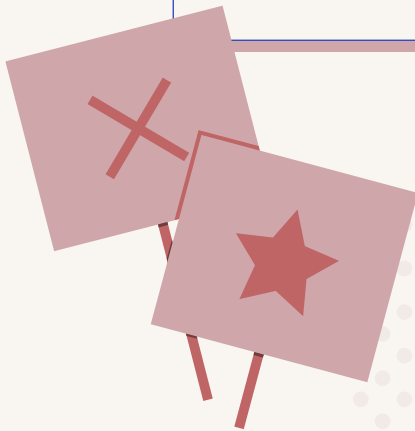
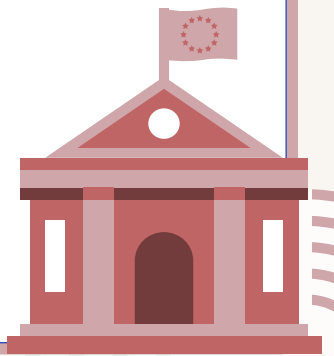
The participants were subsequently asked whether the suggestions the European Parliament has made contribute to the feasibility of complying with the requirements for data quality, both in general and regarding the specific requirements.

In the meantime, the European Parliament has worked on another version, which requires data sets to be up to date, and to the best extent possible, taking into account the state of the art, free of errors, and to be as complete as possible.

Now, we would like you to comment and engage in a discussion of the following aspects:

1 In practical terms: How do you perceive the requirements in Article 10(3)? What are aspects where you see the main challenges?

2 What do you think about the new suggestions from the European parliament (i.e., how feasible are the requirements regarding the data criteria?) Does this solve the challenges that you might have mentioned in point 1?



The respondents are divided on whether the suggestions are improving the proposal. Some feel slight improvements have been made, and that they are more forward looking. Others feel that the requirements are still ambiguous and difficult to measure due to their vagueness, relativity, and subjectivity. A respondent noted:

"In general, requirements are fair and reasonable. However, what is missing is the weighting balancing and risk assessment-based approach here. All these criteria are subjective, relative and not absolute, subject to huge margin of appreciation. Thus, it is easy to manipulate them on one hand, and even more easy to disqualify the data-set as not up-to-date, not sufficiently error free, not representative enough, etc. Thus, all in all, there should be a provision saying that these requirements are subject to proper assessment of (i) cases, where data sets may be subject to less rigid requirements, and (ii) parameters, which are attributable to all these quality features."

Tgs Baltic

Regarding the requirements "error-free" and "complete", respondents noted that the addition to be as complete "as possible" is an improvement as it reduces the burden to some extent.

However, respondents feel that this is still vague because it is unclear what "best as possible" means.



The AIA introduces data quality requirements for providers of high-risk AI systems. Since data is such a central element to many state-of-the-art AI systems, these requirements can have a great impact on organizations developing AI systems. This is reflected in the participants' responses when asked about their general impressions of the data quality requirements. Participants generally seem to agree that the data quality requirements mentioned in the AIA are relevant to consider and they are confident that they can meet (most) of the requirements.

Participants have a high degree of confidence that they can show that the data for their use case is relevant and representative and has the correct statistical properties. However, they do note that it is very hard to quantify these requirements, e.g., how do you measure the relevance of a data set in the context of a particular use case? They remark that these requirements can be highly subjective as well.

Without further guidance, clear and objective methods, and metrics for establishing compliance with these data quality requirements, this provision in the AIA is seen as impractical. This will likely lead to legal uncertainty, which may undermine the development of AI in Europe. Participants also noted that without clear and objective metrics, (objective) enforcement is also impossible.

Finally, the requirements "free of errors" and "complete" are seen as completely unfeasible. Participants observed that data is never complete and never free of errors. The absolute nature of these requirements makes them impossible to comply with in practice. Keeping them in as absolute requirements will make it de facto impossible to provide high-risk AI systems in Europe. The "best effort" requirement that was introduced by the European Parliament is seen as an improvement but still falls short of the mark according to the participants.

Based on the above, we may conclude that while the data requirements are important in ensuring trustworthy AI, the subjective nature of these requirements, and the absoluteness of their wording which makes them almost impossible to achieve, leads to legal uncertainty, which in turn will negatively impact the development of AI in Europe. In their current form, the data requirements may actually act against one of the goals of the AIA – the uptake of AI in Europe.¹⁶

Giving clear and objective metrics for establishing compliance with the data requirements in different contexts, as well as providing guidance on how to ensure the data is relevant and representative and has the appropriate statistical properties, will improve this requirement in the AIA. When it comes to "free of errors" and "complete," the legislator should consider whether these requirements can be rephrased.





Activity 4

Technical documentation

This activity was aimed at gaining a better understanding of whether the participants think they can draw up the technical documentation required under the AIA. To this end, we first asked participants to rate their level of confidence on whether they can provide a general description of several aspects of their AI system (Annex IV (1(a–g))) and where they might have doubts. We asked the participants if they understand the level of detail that they must provide to comply with the technical documentation requirements. We then listed which elements providers should cover in their technical documentation, and asked participants whether it is clear that these aspects should be included in their detailed description of their AI system.

Task 1:
Describing system characteristics

We presented participants with Annex IV of the AIA, which outlines the elements required for the "general description of the AI system":



ANNEX IV

Technical Documentation referred to in Article 11(1)

The technical documentation referred to in Article 11(1) shall contain at least the following information, as applicable to the relevant AI system:

1

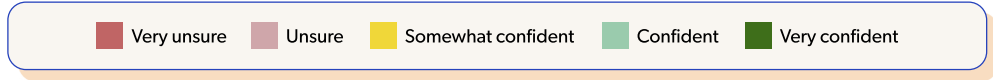
A general description of the AI system including:

- (a) its intended purpose, the person/s developing the system, the date, and the version of the system;
- (b) how the AI system interacts or can be used to interact with hardware or software that is not part of the AI system itself, where applicable;
- (c) the versions of relevant software or firmware and any requirement related to version update;
- (d) the description of all forms in which the AI system is placed on the market or put into service;
- (e) the description of hardware on which the AI system is intended to run;
- (f) where the AI system is a component of products, photographs or illustrations showing external features, marking and internal layout of those products;
- (g) instructions of use for the user and, where applicable installation instructions;

Annex IV(1) AIA

We asked participants to rate their confidence in describing the aspects of their AI system required by the AIA and asked them whether they were unsure about (any) elements.

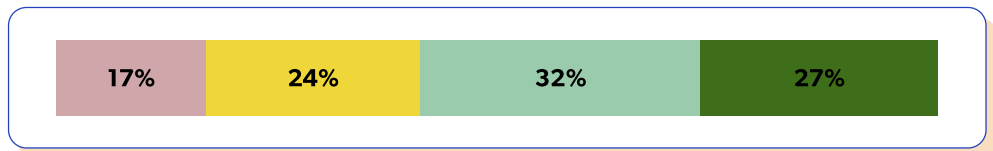
How confident are you that you can provide a general description of your AI system regarding the following aspects:



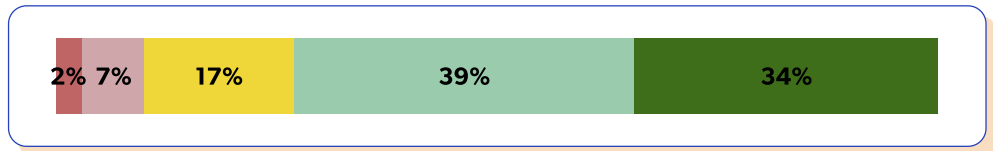
(a) its intended purpose, the person/s developing the system, the date, and the version of the system



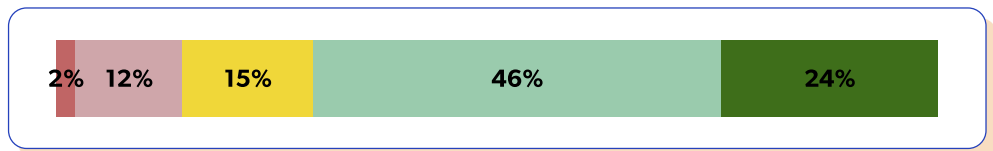
(b) how the AI system interacts or can be used to interact with hardware or software that is not part of the AI system itself, where applicable



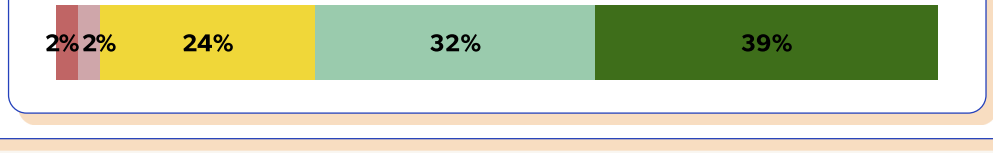
(c) the versions of relevant software or firmware and any requirement related to version update

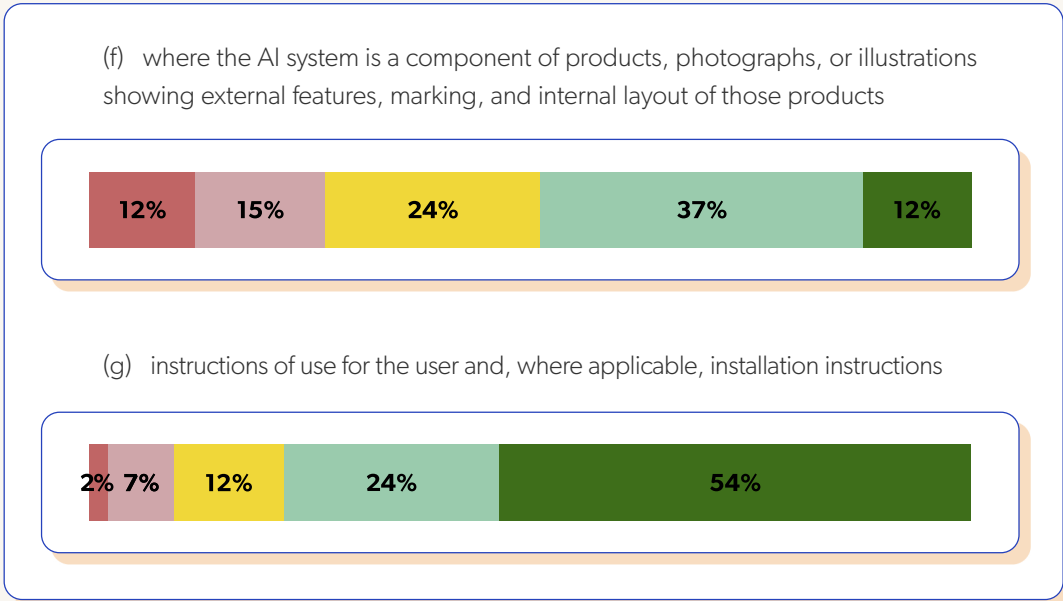


(d) the description of all forms in which the AI system is placed on the market or put into service



(e) the description of hardware on which the AI system is intended to run





How confident are you that you can provide a general description of your AI system regarding the following aspects? [N = 39]

In general, respondents noted the granular level at which technical aspects need to be described. While they feel confident they can accurately describe their systems, they do worry that this will demand an excessive amount of work. Some participants question the need for this requirement, particularly in light of the fact that the regulator cannot do anything with this information without significantly increasing their headcount.

Participants also noted that while the requirements are quite granular, it is unclear how they should be described and in which level of detail. This, once again, may lead to legal uncertainty.

Describing aspects of the system where it interacts with other systems, or where it's a component of other systems, is the most difficult requirement according to the respondents.

Regarding (b) "how the AI system interacts or can be used to interact with hardware or software that is not part of the AI system itself, where applicable," participants were unsure what the scope of "interacts" is. Many AI systems are designed to be embedded into

software, for instance, a website. A participant wonders whether accessing the model through a browser would constitute an "interaction" of the browser and the model.

Regarding (f) "where the AI system is a component of products, photographs, or illustrations showing external features, marking, and internal layout of those products," participants made clear that this is a very obscure requirement. It is intended to target AI systems that are elements of other systems, but respondents noted that this makes little sense to them when it applies to software:

"This is completely obscure. What are 'illustrations showing external features'? What on Earth is meant by 'internal layout'? The AI system produces a score/ranking which is displayed on a dashboard for human project managers to take a decision on; does this make the 'AI system' a 'component of products'?"

Translated

Task 2:
Level of detail
of description

Participants received the section of the policy that outlines aspects of the AI system that require a detailed description of the AI system or of the process for its development. We asked them whether they understood the level of detail that is required by the following clause:



ANNEX IV

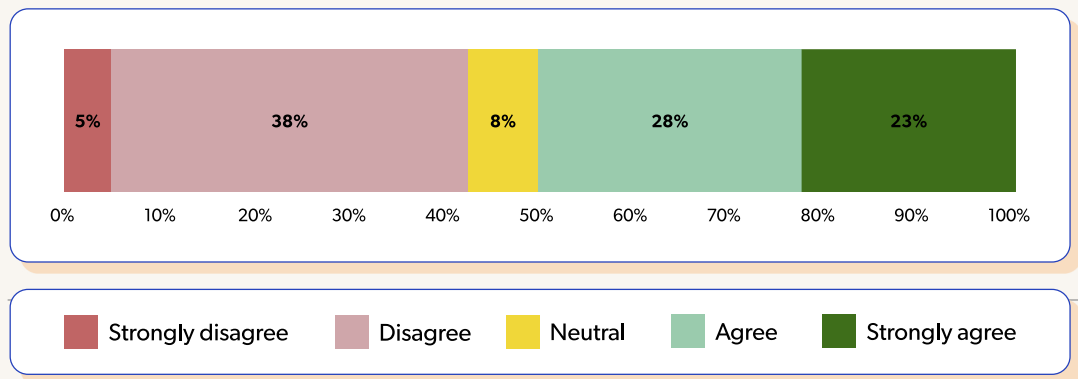
The technical documentation referred to in Article 11(1) shall contain at least the following information, as applicable to the relevant AI system:

- 2** A detailed description of the elements of the AI system and of the process for its development, including:
 - (a) the methods and steps performed for the development of the AI system, including, where relevant, recourse to pre-trained systems or tools provided by third parties and how these have been used, integrated or modified by the provider;
 - (b) how the AI system interacts or can be used to interact with hardware or software that is not part of the AI system itself, where applicable;

Annex IV(2)(a) AIA

Respondents were divided, with half of the participants agreeing that it is clear how much detail the methods and steps performed for

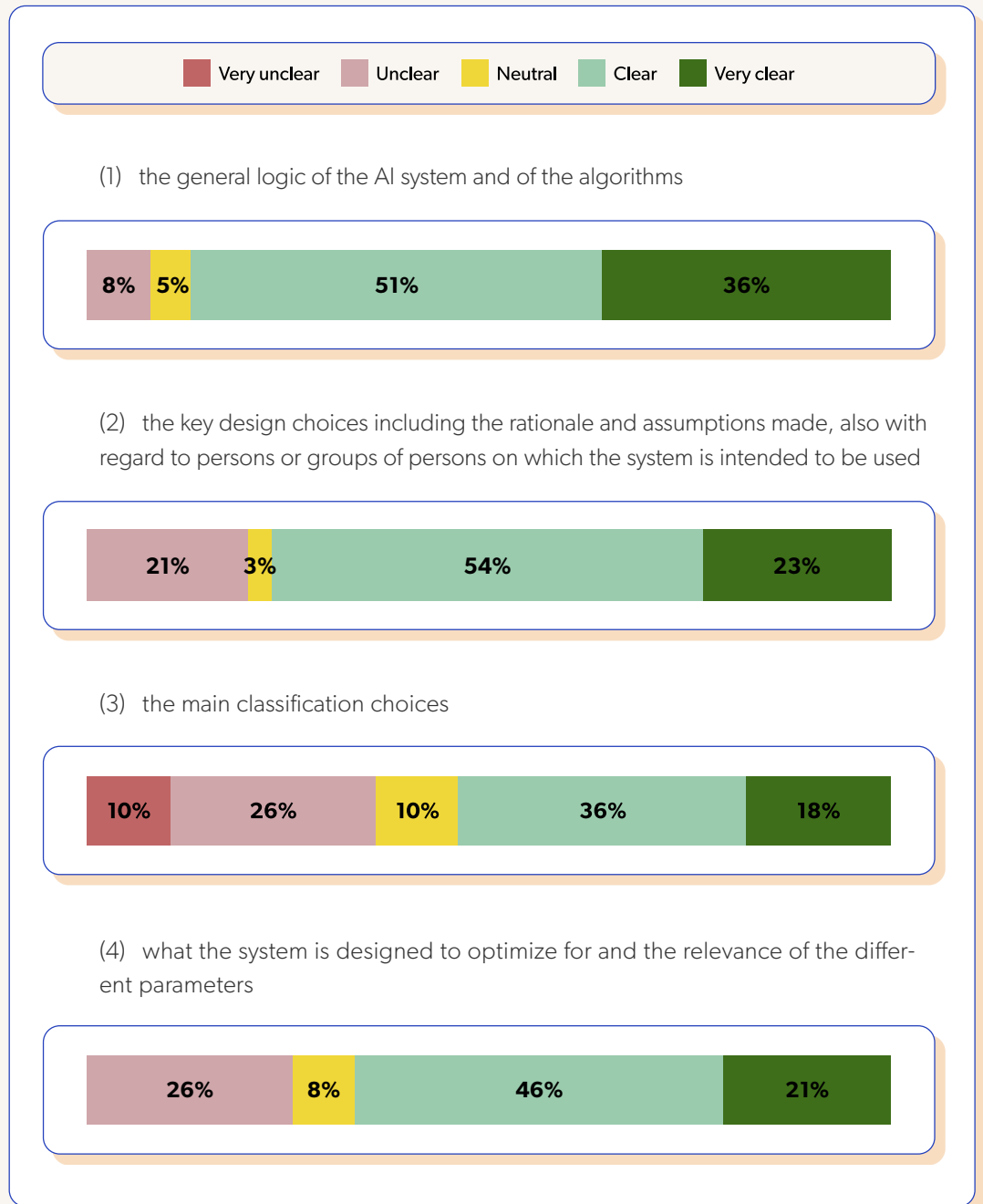
the development of the system need to be described in, while the other half disagreed with that assertion.



I understand the level of detail I must use to describe the methods and steps performed in developing my AI system. [N = 40]

Task 3:
Completeness
of description

To check whether participants understood which elements need to be included in the description, we provided them with a list of elements and asked them whether it was clear that they need to be included.



To what extent is it clear to you what should be included in your AI system's detailed description and the process of its development of the following options? [N = 39]

The responses show that, overall, the requirements are clear, but that a significant minority finds some elements unclear. Providing users with information about the main classification choices is the most unclear element for respondents. More detailed information on what is

being optimized, and the relevance of parameters is also unclear to some of the respondents.

Regarding the key design choices (2), participants noted that the level of detail is not clear. One participant wrote:

“There's just too many things that are intuitively assumed or decided, you cannot document everything because you usually don't know how much you actually assumed. I worked on several AI papers for international conferences and even there in a completely sandbox model it was very hard to list everything you assumed.”

Gossik AG

understand what this meant. Others noted that the classification choices might not be clear before the system is put into service. According to one respondent, unsupervised models are capable of creating their own classification while in service. This means it might be impossible to satisfy this requirement for some systems.

The relevance of parameters (4) suffers from the same issue as mentioned above. Some parameters and their relevance might be unavailable before putting the systems into service, according to the respondents. In addition, a participant noted that deep architectures might create and use parameters in a way that is very hard to evaluate.

The main classification choices (3) are also a challenge. Some participants simply did not

**Task 4:
Information
about functioning
and control**

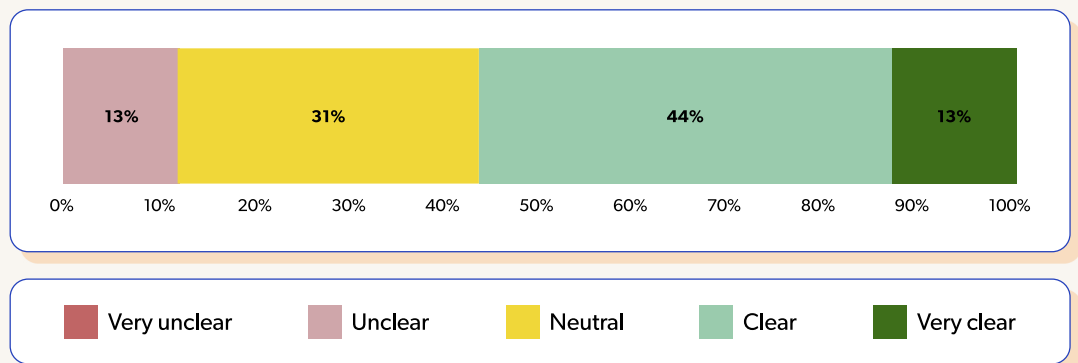
For this task, we provided participants with the following excerpt from the AIA and asked them about its clarity and feasibility.



ANNEX IV

Detailed information about the monitoring, functioning, and control of the AI system, in particular with regard to its capabilities and limitations in performance, including the degrees of accuracy for specific persons or groups of persons on which the system is intended to be used and the overall expected level of accuracy in relation to its intended purpose.

Part of Annex IV(3) AIA



How clear is it for you what you need to describe here? [N = 39]

Most participants find this requirement (somewhat) clear. However, they note several vague concepts in the AIA. For instance, participants are not sure whether "accuracy" is the correct metric in this instance and how it could be measured:

"Accuracy is arguably not the most reasonable word here. Because it may be accurate, but not adequate. Plus, accuracy should be measured in which terms/ figures? All in all, it sounds well written, but very abstract indeed."

Tgs Baltic

The level of segmentation required and the granularity at which "persons or groups" must be captured are also concepts that need to be better defined and understood.

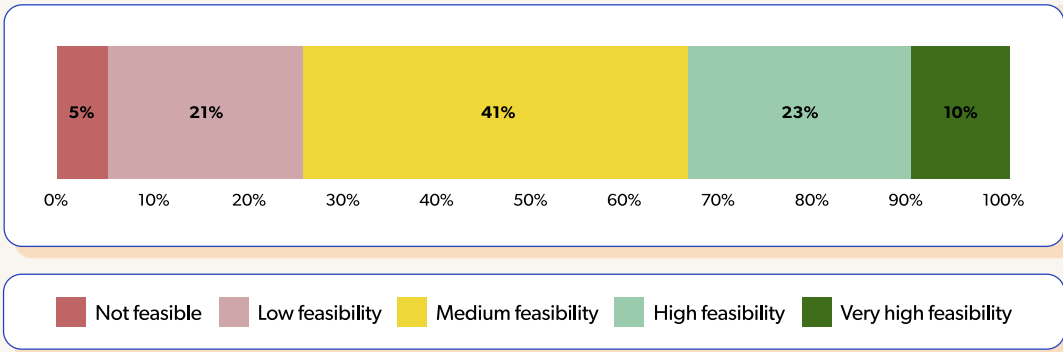
We then asked whether the requirements are feasible to our respondents. The responses

show that the majority is convinced that these requirements are feasible in practice.

When asked what information is challenging to collect in practice, participants shared that the "degree of accuracy" is hard because their systems may be used for many different use cases. Participants also noted that it might be hard to specify "limitations" before placing an AI system on the market, as those might only become apparent when the system is in use:

"A lot of this information may not be available until post model deployment. For example, we may not know the models' limitations until late stages of testing have been completed which involves the deployment of a prototype for end users to interact with."

The VR Hive



From a practical perspective, how feasible is it to give a detailed description of all the required information? [N = 39]

Discussion

The AIA mandates a detailed description regarding the data requirements in terms of (a) datasheets describing the training methodologies, (b) techniques, and (c) the training data sets used, but only "where relevant." We asked the participants in which cases these requirements would be relevant.

Participants note that training methodologies, techniques, and the training data sets are almost always relevant because data information influences the AI system's performance:

"I believe they should be relevant for all systems. I understand there may be a focus on high-risk systems, but why not implement this for all? Having this documentation would assure users data requirements are being properly analyzed."

The VR Hive

"Data requirements are relevant in all cases, particularly when the attribution of liability is so closely tied to transparency throughout the entire ML supply chain/process."

Credo AI

The participants noted that this information is particularly relevant in those situations where there is a risk of biased outcomes or recommendations:

"It would be most relevant where there is a risk of bias in outcomes or recommendations, particularly if the output has a large impact on any person's safety or human rights. Otherwise, it is not as clear when a detailed description of the data is relevant. "

Evo

"Information about training data becomes even more critical when issues are detected with model behavior — for example, when the model is found to make biased predictions against certain groups — because many issues at the model level originate at the data level, and many mitigation techniques for issues like unintended bias involve correcting those issues in the training data."

Credo AI

"It is always relevant to provide information about data requirements of an AI system, because information about the data used to train a system and information about the data that system will need in production to make accurate predictions are absolutely critical in order to understand whether that system will behave as expected — whether that system will perform in the desired way."

Credo AI

A specific issue is the use of pre-trained models. When a pre-trained model is used, the provider does not have access to the information on the training methodologies, techniques, and training data:

"This may get a bit complicated when using pre-trained models [...]. If you are providing a model developed by a third-party, or you are using multiple models with a mixture of your own custom model or pre-trained models in the one larger system, it will be extremely difficult to assess who is responsible for preparing this documentation, and if the requirements for each model need to be compiled separately or all clustered together as they are part of a bigger system."

The VR Hive

So while relevant, the provider does not have access to this information. This raises the question whether the upstream provider should provide this information. This specific issue also underscores the limitation in the taxonomy of AI actors and the responsibility of actors as discussed in Activity 1.



Observations



Participants are confident that they can accurately describe their systems but are worried that it will demand an excessive amount of work. While a significant compliance burden may be justified to avoid the risks associated with high-risk AI systems, it is important to note that this requirement may become a "paper tiger". If regulators do not have enough qualified staff to actually assess the technical documentation, it is questionable whether this requirement will have any effect in practice. It could be that the resources needed to draw up technical documentation could be used in other ways to minimize risk. In particular, at startups and SMEs that have a limited "compliance budget", using these financial resources to draw up technical documentation might not be the most effective allocation of this budget.

Participants also noted that while the requirements are quite granular, they do not contain clear descriptions on how to document these requirements (e.g., level of detail, metrics and methodology). The responses collected from the participants show that there is a level of uncertainty and an operational gap on how participants should comply with some of the requirements listed in the draft proposal, specifically in terms of documenting the characteristics of their AI systems. This makes us tentatively conclude that the high degree of prescriptiveness of the AIA proposal may curtail the level of discretion needed to fulfill its requirements. In fact, by listing a multitude of specific requirements, highly prescriptive laws such as the AIA often end up also requiring additional prescriptive guidance, which can make them more difficult to comply with, as there is less flexibility. While the AIA improves legal certainty by making it clearer what is expected of providers, it unintendedly poses additional challenges to AI companies when it comes to interpreting and complying with such legal requirements. As a result, although well-intended, the effect of providing legal certainty through a high level of prescriptiveness ends up disrupting the balance between high-level norms and the space for legal interpretation and actual compliance. In other words, because the AIA dictates more requirements, additional questions are raised on all these requirements. This contrasts with non-prescriptive laws that have a high level of abstraction, where more is left to interpretation in practice (e.g., through guidance of the regulator, creation of market standards, and/or jurisprudence). Given the high level of detail in the AIA, further guidance by the legislator or the regulator on how to describe their AI systems is desired by the participants.

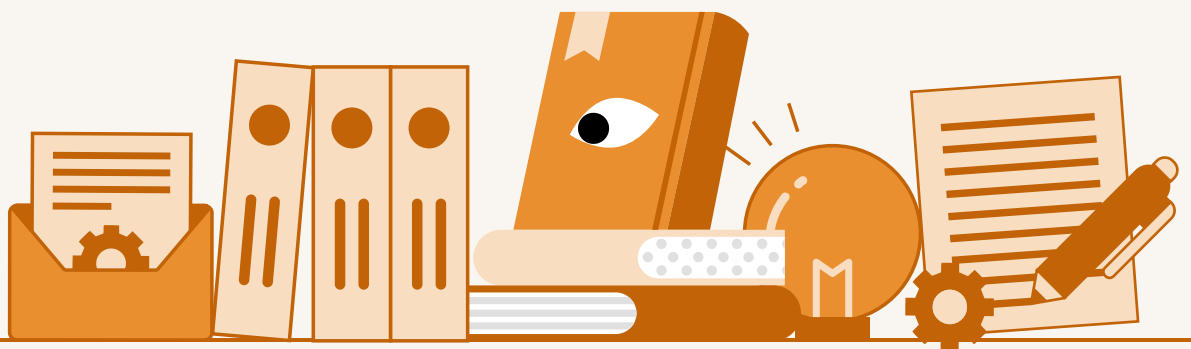
Another important point made by the participants was that it is hard and, in some cases, even impossible to provide all the required information before the AI system is put into production (and thus is placed on the EU market). For instance, unsupervised models are capable of creating their own classification while in service. This means it might be impossible to satisfy the requirement to note the main classification choices.

Looking at the responses of the participants, we get the feeling that unlike more traditional products such as electrical equipment, it seems that AI systems are much more complex and dynamic. Not only does this make drawing up technical documentation more difficult, but it is also more difficult to specify everything before the AI systems are tested in a real production environment.

To a certain extent, the AI regulatory sandbox could address this latter issue. The AI regulatory sandbox has quite strict access requirements though, and given that it must be administered by the regulator, it will also have limited space for participants. It might therefore be interesting to explore the creation of a regulatory environment where AI providers can test their system in a real-life setting with the goal of allowing them to meet the requirements of the AIA, without that meaning that the system is placed on the EU market (e.g., in a pilot or test setup).¹⁷ This regulatory space/environment would thus be in addition to the regulatory sandbox, which is currently focused on the development, testing, and validation of innovative AI systems for a limited time. Of course there should

be sufficient protection against any negative effects through mechanisms in the AIA, but maybe some requirements could be temporarily lifted.

In sum, we can say that while the technical documentation might contribute to the goal of creating trustworthy AI, its operationalization needs further attention. In its current form, the requirement of setting up technical documentation might weigh heavily on companies' compliance budget, diverting attention away from other risk-reducing measures.¹⁸





Activity 5
**Assuring transparency
and human oversight**

The aim of this activity was to assess the level of technical skill that providers of AI systems think users must have and to determine if participants were able to specify what level of technical skill is required to meet the requirement of enabling "human oversight." We hypothesize that the goal of the legislator is to ensure that humans can always assess how an AI system is operating by interpreting its outputs. This presupposes that the human responsible for having oversight has a particular level of skill. We quizzed the participants on the level of skill required and the ability to provide this oversight.

**Task 1:
Designing for
transparency**

We provided participants with the first clause of Article 13 AIA. We asked respondents how confident they felt that they could comply with the requirement of developing a sufficiently transparent system that enables users to interpret outputs and facilitate appropriate use, as described in the following section.



Article 13

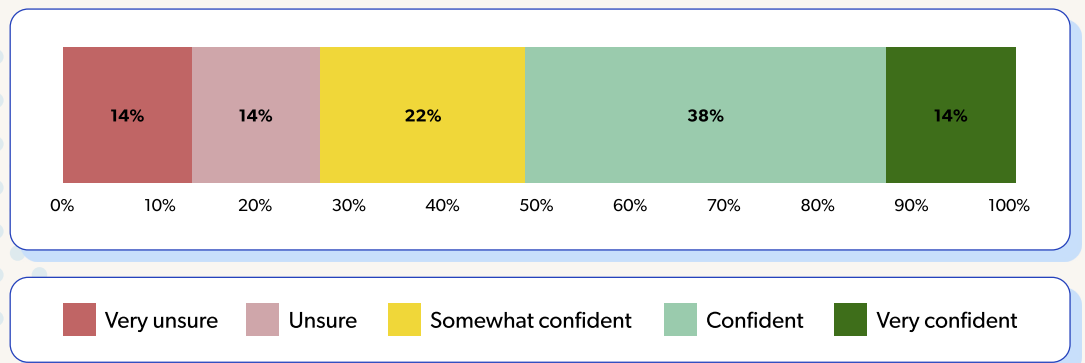
Transparency and provision of information to users

1

High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured, with a view to achieving compliance with the relevant obligations of the user and of the provider set out in Chapter 3 of this Title.

Article 13(1) AIA

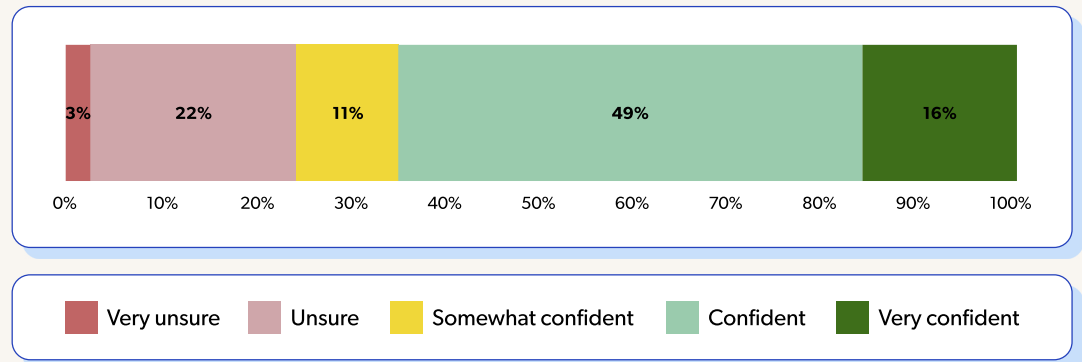
Most respondents are either confident or somewhat confident that they can do this; a substantial minority, however, is either unsure or very unsure about this. In general, respondents concerned about the level of expertise of the users, about the complexity of outputs being too difficult to explain, and about the lack of standardization of evaluation of the transparency of an AI system.



How confident are you that you can design and develop your AI system in such a way that the operation is sufficiently transparent to enable users to interpret the system's output? [N = 37]

When asked whether participants are confident that they can design a system that is transparent enough to use it appropriately, the level of confidence increases. The dif-

ference here is that the design of the system is not so much focused on the explanation, but rather on the ability for the operator to use it effectively.



How confident are you that you can design and develop your AI system in such a way that its operation is transparent enough for users to use it appropriately? [N = 37]

While participants were generally confident that they could build such a system, they did note some challenges such as complexity, trade-offs between explainability and performance, users not following instructions, gaming the system, and IP protection concerns.

A first issue is the complexity of models and the ability to effectively provide relevant information to the end user (the operator) in an understandable and easily digestible format:

“Some models are very complex and it takes time to develop and understand those. I don’t have a clear idea how to pass this knowledge to the end user in digestible format.”

DLabs.AI

The second issue is that of performance. Using a model that is more transparent might come at the cost of performance. While a simple, understandable model might be more transparent, a black box model (e.g., a deep neural net) might perform much better. Without implementing explainable AI solutions, there will be a trade-off between transparency and performance:

“You can select model approaches that are more explainable; however, for some applications, these will be significantly less effective/performant than unexplainable methods. So the question becomes what trade-offs you want and need to make for a given application; is explainability always critical, or are there some applications where performance is more important?”

Credo AI

A third issue that was noted was that users might not always follow instructions and any information provided, or they might ignore them altogether, or interpret explanations incorrectly:

“We help users understand how they should interact with our system, but they could ignore it and interact in their own way.”

indigo.ai

The fact that AIA requires users to act in accordance with the instructions of the provider in Article 29(1) might mitigate the risk of users arbitrarily deciding not to follow instructions, as acting against the instructions will likely shift liability from the provider to the user.

Also, questions were raised with regard to the recipient of the information and the level of detail required in the disclosure. Some participants worry that disclosing information to parties other than the user (similarly to the requirement under article 22 GDPR which mandates data controllers to disclose the logic of the automated decision-making process to the data subject) might enable these parties to "game the system":

"You might not want to disclose exactly how AI is working in order to prevent assessed companies from gaming the system. Not sure if this would be prohibited by that."

Kodex AI


Finally, participants noted intellectual property concerns: making the operation of the model transparent/explainable might entail divulging details about the system that are protected by intellectual property rights.

Task 2:
Human oversight

The skill level of operators also plays a role in enabling effective human oversight. We provided participants with the following section from the article on human oversight:

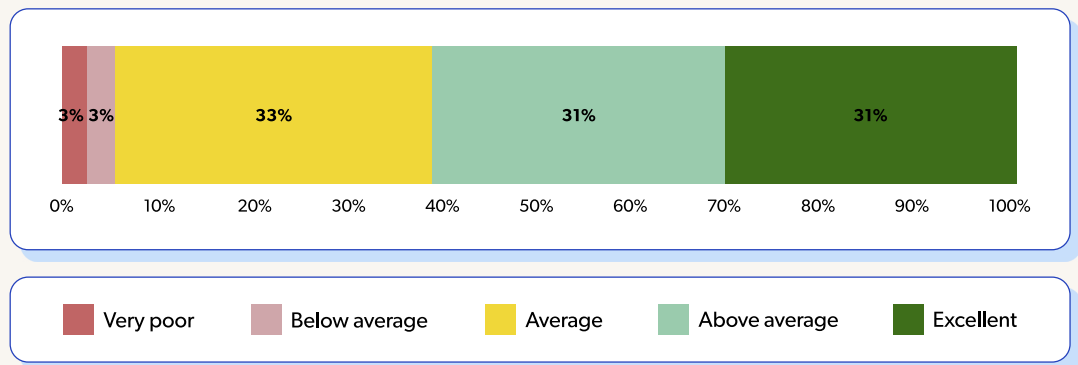
Article 14

"High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use."



Article 14 (1) AIA

We then asked the participants about their expectation of the required level of skill a person tasked to oversee a high-risk AI system must have. Responses show clearly that providers expect that operators need to have (higher than) average skill level:



What is your expectation of the level of technical skills required to oversee a high-risk AI system? [N = 36]

We feel that this is a relevant insight for EU AI policy in general and the AIA in particular. There is a shortage of qualified AI workers in the EU. When we look at the answers of the participants here, we see that the requirement

for human oversight will require highly skilled workers. For this requirement to be effective (and indeed the entire AIA), it is important that the pool of workers with (technical) skills in the AI field grows.

Discussing AI systems’ design, interpretation, use, and risk management

In this activity, we asked the participants about the level of technical skills that users need to have to 1) understand the design of the AI system, 2) interpret the system’s output, 3) use it appropriately, and 4) mitigate health and safety risks.

1) Level of technical skills that users need to have in order to understand the design of an AI system in general

The common view is that an AI system should be designed in a way that users (operators) need no or little technical skills to interact with and make use of such a system, as knowing the inner functioning of a model could also undermine its commercial value.

"Users do not need to know technical details at all. It's on the application to provide an explanation (to be understood by humans) of taken decisions and their motivations."

Manent.AI

"I don't know if our users need to understand the design of our AI systems. That would seriously limit the commercialization!"

Voxist

What we observe here is that the AIA does not make a clear distinction between different audiences. For instance, within the user (which will generally be a company or a government), there are different roles when it comes to using

and assessing AI. Participants seem focused here on the person who uses the system on a day-to-day basis (what we call "the operator"). This could be, for instance, a doctor or a judge. Within the user there might also be a function that is more focused on managing the AI systems used in the company (e.g., an AI portfolio manager or an AI risk manager). These are different audiences with different levels of (required) skill when it comes to interpreting AI outputs.

The participants suggest some basic aspects that users (those using/operating an AI system) and subjects (those subjected to AI decision-making) might want to know:

- That they are interacting with an AI system
- How AI works in general
- What the intended purpose of the AI system is
- The improvement on the outcome of using AI
- The role of data sets
- Why biases occur

Fortunately, most of these elements are covered in the AIA. However, the target audience for this information may be different (e.g., a regulator vs. a user) than is currently specified in the AIA. When it comes to providing instructions to the user, it might be useful to include the elements mentioned above.¹⁹



2) Level of technical skills that users need to have in order to interpret the system's output

Some of the participants stated that only a low level of technical skill should be required to interpret the output of an AI system. If this is not the case, it would make it impossible to use AI systems in practice, as they could only be operated by AI experts rather than normal users that rely on these systems (e.g., doctors or judges).

Some of the respondents mentioned different audiences for explanations/instructions:

"I think the interpretation can be done on multiple levels and for multiple stakeholders. If the system is built in clever ways, all type of stakeholders can be included. If you buy a airplane ticket today, you don't need to check the certification documents for the engines in the airplane, because you know that the airline is certified."

2021.ai

Other participants pointed out that the level of technical skill required depends on the complexity of the AI system. Furthermore, the level of skill depends on factors such as the sector in which the system is used, features of the system, and intended use of the system.

"The required level of user expertise would be a matter of numerous factors, including the application domain, the system's design complexity (partially a matter of the inputs/outputs number and types), the outputs' semantic complexness, the quality of the manual/tutorial explaining the health and safety risks, the quality of the manual/tutorial explaining to the user what a biased AI output may look like (a few indicative examples would suffice I think, as an exhaustive list would be very difficult to produce, if not unrealistic), and the quality of the manual/tutorial explaining the risks of the human biased interpretation of the AI results (as humans, we have also our biases)."

Future Intelligence

3) Level of technical skills that users need to have in order to use it appropriately

Participants suggested that a low level of skill is needed to use the AI system appropriately. However, several participants added that high-risk AI systems require training on how to use them, for instance, via documents explaining the functioning of the AI system.

"Appropriate use and caveats should be explained clearly in a document detailing the functioning of the AI method. Again, the document might be more or less technical depending on the expected final user."

MLcube

"One could imagine that 'high-risk' AI systems should require users to have some training on how to properly use the system and mitigate risks (similar to how someone needs to pass a driving test in order to legally operate a car). The required knowledge here doesn't necessarily need to be technical (again, most driving tests do not ask drivers to describe the components of a combustion engine)."

Credo AI

This matches with the requirements in the AIA that require providers to provide adequate instructions to the user on the use of the AI system and the interpretation of its outputs.

4) Level of technical skills that users need to have in order to mitigate health and safety risks

The answers here are similar to the answers above. Some participants believe that an AI system should be built so that a low level of technical understanding is needed. Others fa-

vor an average level of technical skill and think that a higher level is needed when the system and its output is more complex. These differences in perspective might be explained by the types of AI systems the participants provide (i.e., more complex vs. less complex, or used in a high-risk environment vs. a low-risk environment).

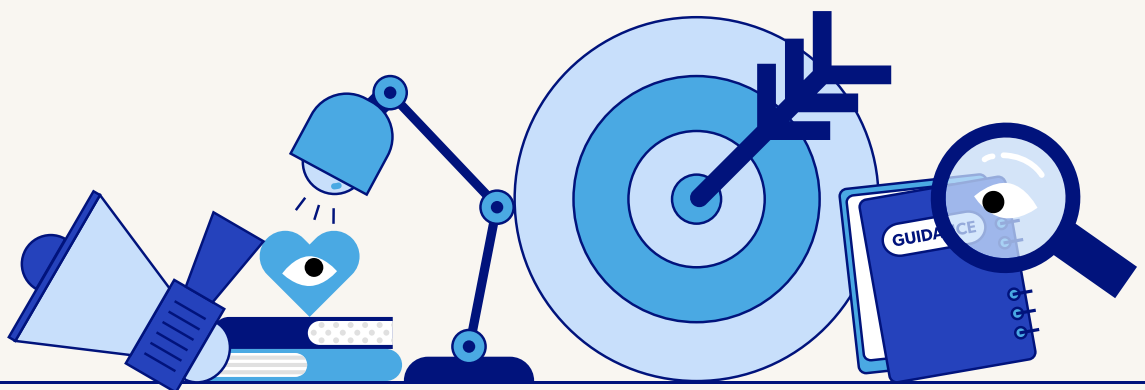


Most participants argued that AI systems should be designed in such a way that users with no or little technical skill or understanding can effectively interact with and make use of AI systems. This is regardless of whether users are required to interpret the system's output or use the system appropriately. If this is not the case, then it could seriously hamper the ability to provide and use AI systems in Europe, as the level of skill necessary for operating the system would be too high.

The participants therefore make a distinction between the operation of an AI system and its oversight. Oversight requires a different skill set (and thus likely a different role) than using the system in practice. This implies that different types of information, explanations, and instructions are necessary for these different target groups. For instance, an AI risk manager tasked with human oversight might need different information than an end user operating the AI system (e.g., a doctor or judge using a decision support system). Participants also agree on the fact that effective human oversight will require a much higher level of skill than operating an AI system.

The AIA does not (clearly) differentiate between these different roles. In Article 14(4), it mentions how "the individuals responsible for human oversight" should be enabled by the provider to execute their oversight, but there is no clear distinction between a person relying on the outputs of an AI system versus the person responsible for monitoring its performance. We feel that both these functions have a role to play when it comes to human oversight, but they require different types of information to enable their oversight. For instance, the operator (e.g., a doctor) might benefit more from an explanation of an individual model outcome, whereas an AI risk manager might benefit more from an explanation on the accuracy of a system, bias in the data, etc.

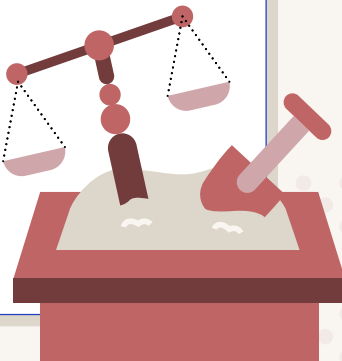
It might thus be beneficial to make the different target audiences for explanations and human oversight more clear and explicit in the AIA, and provide further guidance and/or requirements on how to address these audiences based on their role, technical expertise, and subject matter expertise. This could be done, for instance, through guidelines and/or standardization.





Bonus activity
Regulatory sandboxes

We ended the OLF with a bonus activity on regulatory sandboxes. We asked participants their opinion on the concept of AI regulatory sandboxes as described in Article 53 AIA. More specifically, we asked the following two questions:



- 1 Could you imagine participating in/using such a regulatory sandbox? Please briefly explain why this could be interesting to you.
- 2 One of the goals of the regulatory sandbox is to foster innovation. In your opinion, what are crucial elements of a regulatory sandbox for you to be able to innovate? Please describe and discuss this with others.

Almost all of the participants remarked that a sandbox environment could contribute to more responsible AI innovation and expressed their willingness and even desire to participate in a regulatory sandbox. The ability to test their AI systems in a real-life setting or a close to real-life setting and, in this way, foster innovation was the most important reason to participate in a sandboxing exercise:

“Only in a controlled environment we can foster innovation and allow innovation to find the standard for the future.”

Invrision

“We could imagine ourselves participating in one of these sandboxes, especially around specific AI use cases in for example medtech or fintech. The control mechanisms around AI can (in our opinion) best be tested while shadow running in sandbox environments, in order to test in a simulated, close-to-real-life setting how these mechanisms work. We believe that feedback leads to better models, but also better, more complete control tools and hence foster innovation.”

Deploy

Another reason that was mentioned was the possibility to collaborate with regulators, ensure compliance, and contribute to the operationalization of technical requirements:

“I definitely see the value in having a regulatory sandbox. It helps companies to proactively comply with the intent behind regulations with the support of others (...).”

Evo

"(...) the vast majority of responsible AI initiatives are too high level to be effectively implemented by AI/ML development teams without significant 'translation' or interpretation. We'd like to change that; demonstrating how translation and interpretation happens in 'the real world' through a regulatory sandbox is one tremendously effective way to do this."

Credo AI

Participants also referenced the opportunity that regulatory sandboxes present not only to foster innovation but also to improve regulation. There is an opportunity for this type of experimentation to test both the technical and regulatory systems. This is an important element (and beneficial use case) of regulatory sandboxes that is not included in the current article and its associated recitals (in particular recital 72). By better understanding how the normative requirements of the AIA apply in the real world, one can identify how to improve them:

"Since the repercussions of these systems on the life of EU citizens are so complex and might be understood only after a testing period, I would imagine that it is of general interest to have a preliminary period in which the regulations are evaluated to understand if they are too strict or too loose."

"For sure the regulatory sandbox might be useful to define a final regulatory framework but I do not think this goal is central to this tool."

MLcube

This opportunity is framed by one of the participants as a way to understand how innovation and regulation interact with one another:

"I do imagine ourselves participating in one, yes, and indeed we have done so in the past. For me the value of it is stress-testing our own assumptions and our innovation process in collaboration with regulators, so that both parties (company and regulators) can better identify how innovation and regulation may support and/or hinder one another."

The Newsroom

This is actually aligned with the EP's amendment suggestions for the section on regulatory sandboxes. The co-rapporteur of the proposal, MEP Tudorache, stated that sandboxes would not only be places to test products and comply with regulation, but also be a place for interaction between companies and regulators.²⁰ The IMCO-LIBE amendments presented by Tudorache could achieve such interaction. In fact, a new proposed paragraph of Article 53 states that "[r]egulatory sandboxes shall allow and facilitate the testing of possible adaptations of the regulatory framework governing artificial intelligence in order to enhance innovation or reduce compliance costs."²¹ The rapporteur also calls for the establishment of regulatory sandboxes at local, regional, national, and EU levels. Such amendments list a set of objectives for the sandboxes, which include fostering innovation, facilitating and accelerating access to AI to SMEs and start-ups, improving legal certainty by collaborating with authorities, and contributing to implementing the regulation and to the development of harmonized standards and common specifications.

And as the same participant notes, regulatory sandboxes that are aimed at fostering technical innovation and improving regulation present a number of important benefits:

"(...) This can 1) help us better understand why certain legislation is in place, making it feel less like a burden and more like a boundary that is necessary and good; 2) better inform policy, as it can surface unintended positive and negative consequences; and 3) help us better innovate, as good boundaries are a useful ingredient for responsible innovation."

The Newsroom

While users were overwhelmingly positive about the concept of a regulatory sandbox, there were also concerns that individuals and society may be negatively affected if the sandbox is not properly deployed.

The participants listed the following crucial elements that a sandbox must contain in order to foster innovation:

- collaboration,
- transparency,
- guidance and legal certainty, and
- protection from enforcement.



Collaboration

A first element that was mentioned was collaboration with other participants and the regulator. In order for a sandbox to effectively foster innovation, it is important that the sandbox provides a learning environment for both the participants and the regulator:

"(...) Innovation requires openness. I think the most important element of any regulatory sandbox would be a collaborative environment where people share their own problems and ideas to iterate on the best strategies and improve them in cycles, over time."

Evo

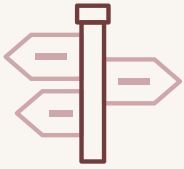


Transparency

A second element that was mentioned was transparency. Transparency is required to promote trust in the regulatory sandbox. The intentions of the regulator and the participants must be clear and explicit.

"(...) In my opinion, a regulatory sandbox is effective in fostering innovation when designed with transparency and collaboration in mind, in order to provide a trustworthy environment for all the actors."

ASC27



Guidance and legal certainty

Participants described the value of guidance from the regulator, in particular for the legal certainty it can provide:

"What is needed to innovate in a sandbox? Legal certainty. Access to regulatory advice, ideally in the form of 'advance rulings', e.g., committing opinions from the regulator on precisely written questions."

Translated

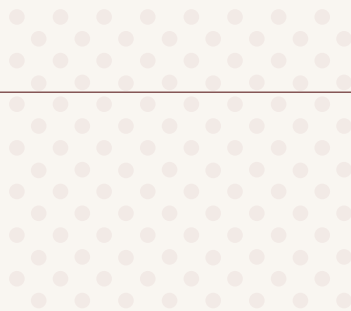


Protection from enforcement

The previous point on guidance and legal certainty ties in with the point of protection from enforcement. Participants note that it is important that a sandbox is a "safe environment" where participants do not have to fear enforcement from the regulator.

"(...) In order for a regulatory sandbox to foster innovation, we believe that it would be beneficial to allow organizations and individuals to participate without fear of retribution — sandboxes need to be a safe space where ideas and issues can be explored without judgment from regulators. Participants need to be able to raise concerns without worrying whether those concerns will be used against them or their organization in the future."

Credo AI



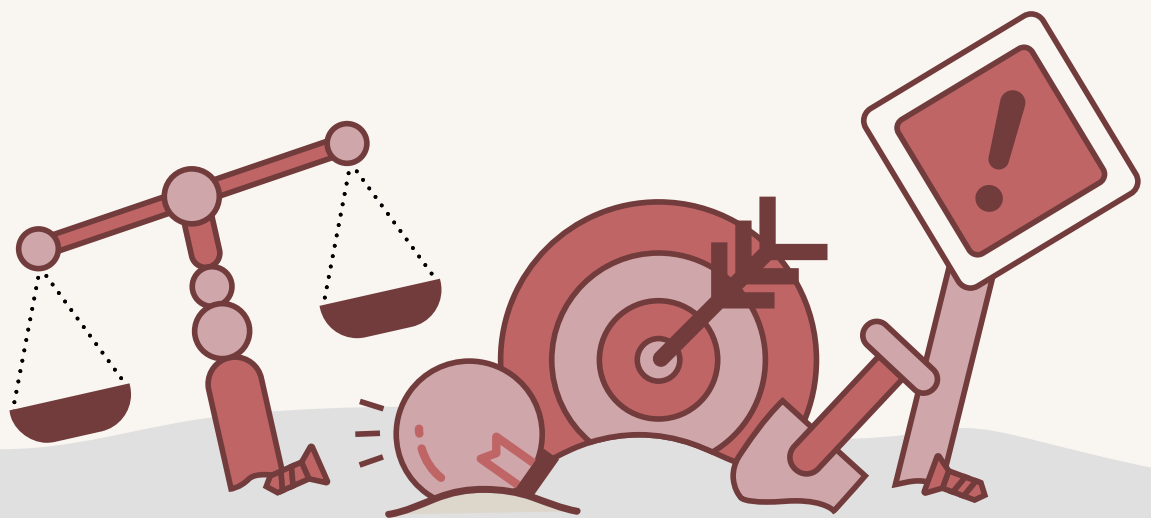


Observations

What we can learn from the reactions of the participants is that the regulatory sandbox provision in the AIA can be an important mechanism not only to foster innovation but also to strengthen compliance and improve the regulatory framework once enacted.

In the OLF, participants remarked that the assessment of risk and technical requirements such as accuracy and performance can sometimes only be established in a real-life setting. This in turn generates a loophole by which it may be impossible for providers to assess the risk of their system (i.e., to comply with the AIA) unless they place it on the market first (i.e., in a real-life setting). The regulatory sandbox may provide a mechanism that would allow for real-life testing on a limited scale, fostering innovation, and strengthening compliance, while providing inputs to improving the regulatory framework in place.

There are, however, several prerequisites that must be met for a sandbox to be effective such as legal certainty and a collaborative environment. These are important elements that are not covered (in depth) in Article 53 AIA. Through implementing acts and guidance at the sandbox level, these elements could be covered.²²





Conclusion and recommendations

Through the OLF we have gathered valuable feedback from participating companies. The participants gave us direct feedback on the AIA based on their expertise in the field of AI, and by interacting with its provisions, they gave us insights into how the AIA might be interpreted once it is enacted.

The overall picture is that most of the provisions in the AIA are clear and feasible and may contribute to the overall goal of creating trustworthy AI. However, there are several areas in the AIA with room for improvement and some provisions that might even undermine another goal of the legislator: the uptake of AI in Europe.

A first conclusion from the OLF is that the binary distinction between the provider and the user which the AIA adopts does not accurately reflect the complexity of the AI ecosystem. Relevant elements for the legislator to consider are: 1) how to deal with "co-production," whereby the provider and the user both have a significant influence over the final product; and 2) the use of pre-trained models and other components in high-risk AI systems. Some of these limitations are being addressed by the amendments of the European Parliament (see, for instance, the Opinion of the JURI committee), which calls for a more granular taxonomy and a clearer allocation of responsibilities along the AI value chain.

A second conclusion is that while participants are able to assess the risks of their AI systems, they seem to focus more on the root cause of the risk (e.g., biased data and model drift) than on the impact the system might have. Based on this observation, we theorize that those subjected to the AIA (providers and users) might benefit from more concrete guidance on risk and risk assessment in the AIA itself, or in subordinate legislation and/or soft law instruments.

An important conclusion from the OLF is that data quality requirements are formulated in such a way that their interpretation is subjective. While this is not uncommon in legislation, it leads to legal uncertainty. Furthermore, the

data quality requirements "free of error" and "complete" are unrealistic and unfeasible according to the participants, particularly in the context of training models, as data sets are hardly ever free of error or complete. According to the participants, the absolute nature of these requirements could impede the uptake of AI in Europe. The current discussion at the European Parliament level might help reduce the limitations of such requirements.

The participants feel that drawing up technical documentation is a significant burden on them. Furthermore, they argue that it will also be a burden for the regulator and other authorities to review the technical documentation. While the participants feel that drawing up the technical documentation is generally feasible, they do argue that they need more clear guidance, templates, and metrics to draw up the technical documentation. Therefore, despite the high level of prescriptiveness in the annexes to the AIA (or maybe because of the high level of prescriptiveness), participants struggle with this requirement.

The questions and discussion on human oversight reveal that for effective human oversight an average or above average level of technical skill is necessary. This implies that significant investments need to be made in skilled labor, both at the side of the user and the regulator. Furthermore, it is not always clear for participants who the audience is for an explanation. Different audiences may benefit from different (types of) explanations.

Finally, what we can learn from the reactions of the participants is that the regulatory sandbox provision in the AIA can be an important mechanism to foster innovation, strengthen compliance, and improve the regulatory framework once enacted. There are, however, several prerequisites that must be met for a sandbox to be effective, such as legal certainty and a collaborative environment. These are important elements that are not covered (in depth) in Article 53 AIA. Through implementing acts and guidance at the sandbox level, there is ample opportunity to address these elements.

Recommendations

While much more research is needed and the scope of this OLF exercise is limited, we tentatively formulate the following recommendations for improving the AIA and its effectiveness in reaching the goals of the legislator:



Consider **revising/expanding the taxonomy of AI actors** in Article 3 and/or more accurately describe possible interactions between actors (e.g., co-production of AI systems and use of open-source tooling) to more accurately reflect the granularity of the AI ecosystem.



Given the difficulty to assess "reasonably foreseeable misuse" (Article 9) and the limited focus on impact of risks, **provide guidance on risks and risk assessment**, in particular for startups and SMEs.



Provide more concrete guidance, methodologies, and/or metrics for assessing the data quality requirements through, e.g., subordinate legislation and/or soft law instruments, standardization, or guidance from the regulator (Article 10).



Revise the data quality requirements "error free" and "complete" as they are considered unrealistic and unfeasible (Article 10).



Provide more concrete guidance, templates, and/or metrics for the technical documentation through, e.g., subordinate legislation and/or soft law instruments, standardization, or guidance from the regulator (Article 11).



Ensure that the requirement for technical documentation does not become a "paper tiger" by ensuring **sufficient qualified staff to actually assess the technical documentation** (Article 11).



Consider **distinguishing more clearly between different audiences** for explanations and other transparency requirements (Articles 13 and 14) in the AIA.

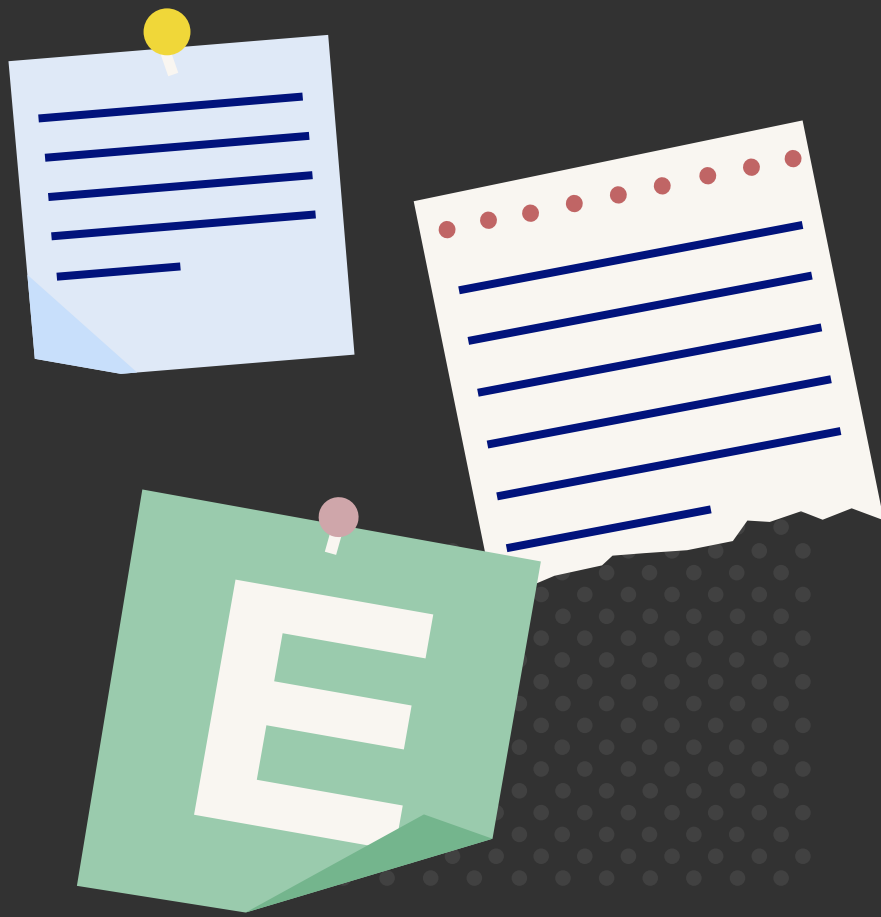


The AIA's success hinges on the ability to execute and enforce the regulation. Therefore, ensure that the future workforce contains **enough qualified workers**, in particular when it comes to **human oversight of AI** (Article 14).



Maximize the potential of regulatory sandboxes to foster innovation, strengthen compliance, and improve regulation. Ensure that, through implementing acts and guidance, **conditions for effective AI regulatory sandboxes are created** (e.g., collaboration, transparency, guidance and legal certainty, and protection from enforcement) (Article 53).





Endnotes

- 1 For more information about the program and the stakeholders involved, see <https://openloop.org/programs/open-loop-eu-ai-act-program/>
- 2 For an applied perspective on how AI explainability can be operationalized based on four different types of audience (among other variables), namely regulator, business partner, consumer, and society, see the Open Loop report on AI Transparency and Explainability. Andrade, Norberto Nuno Gomes de. "AI Transparency and Explainability - A Policy Prototyping Experiment" (2022), at https://openloop.org/wp-content/uploads/2022/07/AI_Transparency_&_Explainability_A_Policy_Prototyping_Experiment.pdf
- 3 Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. COM/2021/206 final. The proposed regulation was presented by the European Commission on April 21, 2021, and is aimed at establishing a harmonized legislative framework for the regulation of artificial intelligence in the European Union's single market. The AIA adopts a risk-based approach to regulate specific uses of AI, thus ensuring adequate levels of safety and fundamental rights protection. See European Commission "A European Approach to Artificial Intelligence", at <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>
- 4 Fuller, Lon Luvois (1964). *The Morality of Law*. New Haven: Yale University Press. p. 91.
- 5 See <https://www.theoryofchange.org/what-is-theory-of-change/>
- 6 Regulatory sandboxes are the theme of a separate pillar of this Open Loop policy prototyping program on the EU AIA.
- 7 We use the term "participants" interchangeably to refer both to the participating companies and their individual representatives.
- 8 See Gils, T., Vranckaert, K. , Benichou, B. "Exploring Policy Prototyping – Some Initial Remarks" (2021), at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3885571
- 9 Note that we did not check for each participant whether they had assigned themselves the correct role under the AIA.
- 10 This can be attributed to our sample selection process, as we specifically enlisted AI companies for this research.
- 11 See European Commission (2022). COM(2022) 496 final - Proposal for a directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive); and European Commission (2022). COM(2022) 495 - Proposal for a directive of the European Parliament and of the Council on liability for defective products.
- 12 Note that participants refer to Aladin both as a "user" and as a "provider."
- 13 "OSS" stands for "open-source software." "Provided on an as-is basis" means that the license might come with a provision saying that the user takes the risks with any issues inherent in using it.
- 14 See Opinion of JURI Committee - Compromise AMs - JURI AI Act - FINAL (30/08/2022).
- 15 A similar observation was made in the context of the 2021 Open Loop policy prototype on Automated Decision Impact Assessment. See <https://openloop.org/programs/ai-impact-assessment-europe/>

- 16 See recital 5 AIA.
- 17 The proposal of testing a product in real-world conditions has been put forward by both the French and Czech Presidencies at the Council level. The Council's compromise text drafted under the French and Czech Presidencies introduce the concept of "testing in real-world conditions," which describes "the temporary testing of an AI system for its intended purpose in real world conditions outside of a laboratory or otherwise simulated environment with a view to gathering reliable and robust data and to assessing and verifying the conformity of the AI system with the requirements of this Regulation." See Council of the European Union, French Presidency Compromise text, consolidated version. Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union - Text de compromis de la présidence - Article 3, paragraphe 1 ter, Articles 4 bis à 4 quater, Annexe VI (3) et (4), considérant 12 bis, at <https://data.consilium.europa.eu/doc/document/ST-9029-2022-INIT/x/pdf>
- 18 See recital 5 AIA.
- 19 See the Open Loop report "AI Transparency and Explainability - A Policy Prototyping Experiment" for a list of different types of target audience for explanations. Andrade, Norberto Nuno Gomes de. "AI Transparency and Explainability - A Policy Prototyping Experiment" (2022), at https://openloop.org/wp-content/uploads/2022/07/AI_Transparency_&_Explainability_A_Policy_Prototyping_Experiment.pdf
- 20 See the European Approach to Regulating Artificial Intelligence with MEP Dragos Tudorache, Co-Rapporteur of the EU AI Act. September 19, 2022. <https://www.csis.org/events/european-approach-regulating-artificial-intelligence-mep-dragos-tudorache-co-rapporteur-eu-ai>
- 21 European Parliament, Committee on the Internal Market and Consumer Protection and Committee on Civil Liberties, Justice and Home Affairs. Draft report (PE731.563v01-00) Harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts Proposal for a regulation (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). June 13, 2022. Amendment 2333 Article 53 – paragraph 5 a (new), at https://www.europarl.europa.eu/doceo/document/CJ40-AM-732840_EN.pdf
- 22 Regulatory sandboxes are the theme of a separate pillar of this Open Loop policy prototyping program on the EU AIA, where we assess whether these instruments as envisaged by the AIA would be attractive to organizations developing AI systems. Incentives and key features of an AI sandbox will be discussed and assessed during a policy jam involving subject matter experts. This will result in recommendations aimed to increase sandboxes' effectiveness and facilitate their uptake.