FROM POLICY TO PRACTICE

PROTOTYPING THE EU AI ACT'S HUMAN OVERSIGHT REQUIREMENTS

Knowledge Centre Data & Society

Artificial Intelligence
Flanders

Citation: Wannes Ooms, Lotte Cools, Thomas Gils and Frederic Heymans (Knowledge Centre Data & Society), "From Policy To Practice: Prototyping The EU AI Act's Human Oversight Requirements", March 2025

Contact: wannes.ooms@kuleuven.be or thomas.gils@kuleuven.be

www.data-en-maatschappij.ai

1. ABOUT THE KNOWLEDGE CENTRE DATA & SOCIETY

The Knowledge Centre Data & Society (KCDS) is the central hub in Flanders for the legal, social and ethical aspects of data-driven and AI applications. The Knowledge Centre brings together knowledge and experience on this topic tailored to industry, policy, civil society and the general public. Specifically, our objectives include:

- **Disseminating information and knowledge** on the ethical, legal and social aspects of data-driven applications and AI. All publications are made publicly available and aim to create a positive and proactive effect between these innovations and our society.
- **Promoting structural initiatives** that strengthen vision development and valorise the social and economic opportunities of data-driven applications and AI among governments, industry and other social actors.
- Stimulating public awareness and debate on the benefits and drawbacks and the social, ethical and legal aspects of data-driven applications and AI, in all layers of society.
- **Building and supporting a network and learning environment** for stakeholders and strengthening collaboration between different policy levels and actors.
- **Contributing to the development of legal frameworks and guidelines** on the use and framing of AI and data-driven applications for policy makers, businesses, organisations and employees. Our policy prototyping project is one of the activities that we develop in order to achieve this objective.

Please visit our website[1] for more information about the KCDS, our objectives and our offering.

[1] https://data-en-maatschappij.ai/en/

2. TABLE OF CONTENTS

3. EXECUTIVE SUMMARY

Article 14 AI Act mandates that high-risk AI systems are designed and developed in such a way that they enable effective human oversight. It establishes the requirement that human oversight be made possible by providers both through the design of the high-risk AI system as well as through organisational measures, identified by the provider and implemented by the deployer. This should ensure that the risks to health, safety and fundamental rights are effectively mitigated.

This project intended to test the human oversight requirements for high-risk AI systems in the AI Act by gathering stakeholders to implement these requirements into prototype compliance documents. We collected feedback on these prototype compliance documents in order to determine best practices and policy recommendations. This report includes the compliance documents drafted by the participants of the policy prototyping workshop in its annex. In addition, the report also features reviewer feedback on those documents and the human oversight requirement contained in article 14. The reviewers who provided feedback on the compliance documents did not necessarily participate in the policy prototyping workshop.

The report starts with an introduction to policy prototyping (part 4) and outlines the course and different phases of this project (part 5). Then, it discusses the prototype compliance documents that were developed and the respective stakeholder feedback (part 7).
The final section contains detailed legal feedback on article 14 AI Act (part 8).

KEY FINDINGS RELATED TO PROTOTYPE COMPLIANCE DOCUMENTS AND HUMAN OVERSIGHT MEASURES

Below, we highlight some of the findings in relation to the prototype compliance documents, based on participant feedback.

Findings on the compliance documents

- **Human oversight governance and role distribution**: a clear human oversight governance structure should be established, containing an allocation of tasks with a clear identification of responsible actors. The compliance documents should describe the profile of these designated individuals, including the expected level of AI literacy. This assessment can be performed by the provider.
- **Tailored output of the AI system**: the terminology used in the AI output and oversight instructions should be tailored to individuals performing human oversight, combining technical and sector-specific terminology for clarity. Technical human oversight measures were considered preferable over organisational measures.
- **Information on risks and user profile**: Correct and complete descriptions of the risks posed by the AI-system and the corresponding human oversight measures, as well as the background of the system (both in terms of intended purpose and use as well as on the technical functioning of the system), were seen as an important element to correctly perform the oversight. A description of the intended user of the AI system was also considered useful for its correct functioning and oversight.

- **Combination with Instructions for Use**: the human oversight requirement must align with article 13 AI Act, which outlines transparency requirements and requires instructions for use. Providers can aid deployers by offering clear, specific instructions for end-users. This approach has proven effective in the first use case; when the human oversight information is structured more like instructions for use, this resulted in (more) comprehensive compliance documentation.
- **Format & language**: the format influenced how 'complete' and 'user-friendly' the reviewers deemed the compliance documents. The documents should have a logical structure with clear language tailored to the target audience (i.e., the individual who needs to perform human oversight). The documents must be drafted in a way that contributes to their concrete use and understanding (visualisation – where relevant – for example generally aids the understandability).

KEY FINDINGS RELATED TO THE HUMAN OVERSIGHT REQUIREMENTS OF THE AI ACT

Article 14 offers broad flexibility for deployers and providers. While beneficial in some regards, such as contextual adaptability, this flexibility makes practical implementation challenging. Sector-specific guidelines, technical standards, and concrete examples are essential to provide clear benchmarks for compliance. Providers otherwise have no way to measure whether they sufficiently comply with the requirements and lack legal certainty. Authorities can also clarify the division of responsibilities for human oversight among the provider and deployer, since this distinction can be vague in practice.

A lack of expertise by the user to correctly perform the oversight and a lack of awareness of the human oversight obligations at both the provider's and deployer's end were considered major concerns for the feasibility of the requirements. Authorities should consider actions to promote this expertise and increase awareness.

Overall, the legal requirements on human oversight were considered desirable to build trust in the AI system. Certain terms in the obligations were considered overly vague and non-concrete leading to a lack of understandability and a need for clarification. Similarly, the proportionality of the measures was difficult to determine and might lead to providers taking the "path of least resistance" without certainty that this leads to sufficient compliance.

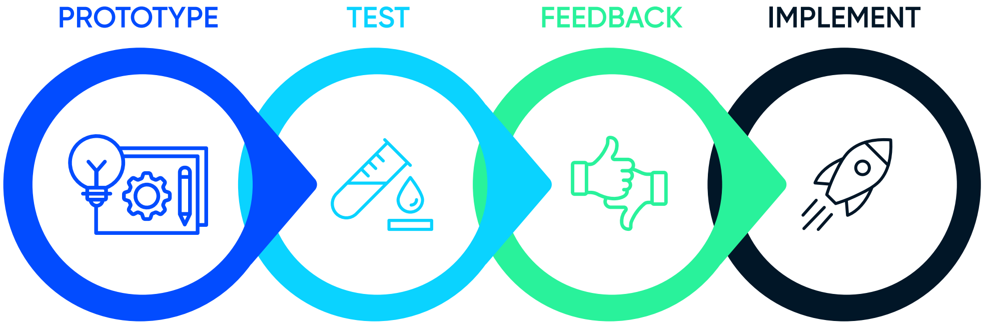
KEY FINDINGS RELATED TO POLICY PROTOTYPING

Throughout the project, the concept of policy prototyping has garnered positive feedback. Many participants recognised the significant value that policy prototyping may bring, thereby emphasising the usefulness of exploring the application of regulatory requirements and provisions on a real use case. Participants agreed that this method can add much value to the policy implementation process, as it turns abstract obligations into a tangible reality. Both the use case providers, who receive input directly applicable to their AI system, and the other participants, who build experience on applying the obligations in accordance with considerations stemming from the use case, gain insightful practical knowledge. The positive reception underscores a broader consensus emphasising the importance of interactively involving a diverse array of stakeholders in the policymaking process. By gathering comprehensive insights from these different perspectives, policymakers can ensure a solid foundation for the policy implementation process.

4. INTRODUCTION

4.1. Introduction to Policy Prototyping

Policy prototyping refers to an alternative way of policymaking, comparable to product or beta testing. It can be understood as a form of user-centred policy design or applying the design thinking methodology to the legislative or policymaking process. Policy prototyping should enable policymakers to map the effects, strengths and limitations of a proposed policy and lead to more effective and evidence-based policymaking while avoiding the societal costs of 'bad policy' negatively impacting stakeholders. Typically, a policy prototyping project consists of multiple phases:



- **Prototype**: prototyping implies the creation of basic models or designs for a machine or other product to test an idea or a concept in practice. In this context, prototyping entails drafting a new policy or law. Such prototypes can be elaborate or minimal, allowing to test specific features and find out 'what works' through several iterations.
- **Test**: a group of stakeholders performs a mock compliance exercise and implements the envisaged legal requirements.
- **Feedback**: participants provide feedback in relation to the mock implementation of the policy prototype.
- **Implement**: this feedback is used to evaluate if the law is effective and 'fit for purpose' and to complete and/or amend it accordingly, issue additional guidance, highlight ambiguities etc.

In summary, policymakers and stakeholders can create tangible and practical prototypes of proposed policies and related compliance documents using this approach. These prototypes allow them to test and refine the policy measures before committing to a full-scale implementation.

Policy prototyping can help identify potential gaps, challenges, or unintended consequences at an early stage of the policymaking process. It enables policymakers to make necessary adjustments and improvements to the policy, and stakeholders to prepare for future policy. In essence, policy prototyping may bridge the divide between policy design and actual implementation, enhancing the effectiveness, feasibility and acceptance of policies while minimising the risk of unanticipated policy mistakes or failures.

At the same time, policy prototyping projects should also consider some possible concerns for which they should ensure transparency or accountability. More specifically, the group of participants involved in a project ideally reflects the diverse group of stakeholders affected by the envisaged policy, while public transparency regarding the participants also needs to be ensured. Additionally, policy prototyping projects will generally be conducted with small testing groups. This may lead to casuistic results, reducing their representativity and scalability, as the results may not be applicable on the large scale on which regulation usually applies. In part 5, we will explain in more detail how we applied this approach (including the concerns) in the policy prototyping project which is the subject of this report.[2]

4.2. Policy Prototyping at the KCDS

Policy prototyping has been used for several years in the work of the KCDS to provide stakeholders with more insight into the application of the EU AI Act.[3]

In 2023, we implemented a policy prototyping project around the AI Act. This project focused on the **EU AI Act's transparency requirements**. More precisely, it concerned the transparency requirements for high-risk AI systems (article 13 AI Act) and the transparency requirements for "certain AI systems", including interactive AI systems and AI-generated/deep fake content (article 50 AI Act). The main findings of this project can be consulted in our report "From Policy to Practice: Prototyping The EU AI Act's Transparency Requirements".[4]

The current project was launched and conducted in 2024 and focuses on article 14 AI Act. In the course of this project, we pursued four objectives:

1. **Examine** the envisaged human oversight requirements in detail;
2. Create **operational guidance** that includes prototype compliance documents for high-risk AI systems (under article 14 AI Act);
3. Gather **feedback** on the human oversight requirements and their applicability, feasibility, desirability and understandability;
4. Provide our **findings and lessons learned** to policymakers and other stakeholders.

[2] For more information on policy prototyping, see: B. Benichou, T. Gils and K. Vranckaert, Design thinking in the legislative process: the key to useable legislation?, April 2021, https://www.law.kuleuven.be/citip/blog/design-thinking-in-the-legislative-process/.
See also: T. Gils, K. Vranckaert and B. Benichou, "Exploring Policy Prototyping – Some Initial Remarks", July 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3885571

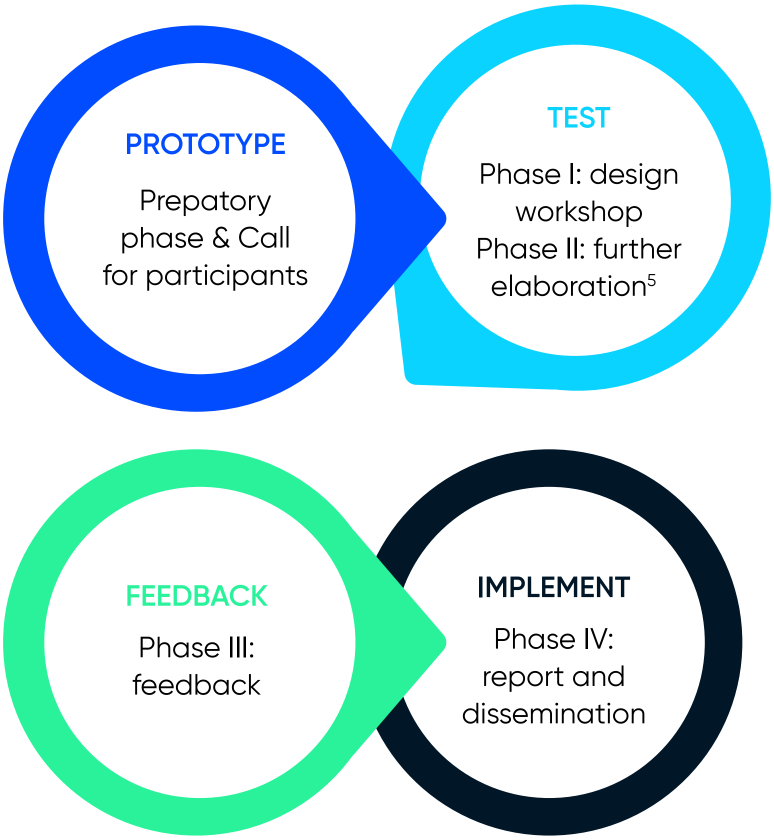
[3] An overview of all initiatives can be found here: https://www.data-en-maatschappij.ai/en/policy-prototyping

[4] T. Gils, F. Heymans and W. Ooms (Knowledge Centre Data & Society), "From Policy to Practice: Prototyping The EU AI Act's Transparency Requirements", January 2024. https://data-en-maatschappij.ai/uploads/Policy-prototping-report-jan2024.pdf

5. POLICY PROTOTYPING: METHODOLOGY AND PROCESS

In this part we will explain the methodology that we followed for this policy prototyping project, which is similar to the previous project. We believe that this is necessary to enable a correct interpretation and use of the results included in this report.

The policy prototyping project outlined in this report was initiated in March 2024, commencing with an initial phase dedicated to the selection of the 'policy prototype' to be tested (i.e. article 14 AI Act). Subsequently, a call for participants was issued, and interested stakeholders were identified. This group was invited to a design workshop in September 2024, during which participants collaborated in small groups to draft mock compliance documents (based on real use cases) as required under the EU AI Act. These documents were further elaborated over autumn 2024. Both the article 14 AI Act as well as the prototype compliance documents were then subject to feedback via qualitative online or in-person interviews. The findings of those interviews are aggregated in this report. The visual below illustrates how our phases map on the (theoretical) phases mentioned in part 4.1.


PROTOTYPE
Prepatory phase & Call for participants
TEST
Phase I: design workshop
Phase II: further elaboration[5]
FEEDBACK
Phase III: feedback
IMPLEMENT
Phase IV: report and dissemination

[5] The design workshop and the further elaboration could also feature in the prototyping phase as we created prototype compliance documents in those phases. However, as our main goal was to test the AI Act's requirements, we decided that they rather fit under the testing phase.

5.1. Preparatory phase: decision on legislative framework and practical considerations

Our choice of article 14 AI Act for this policy prototyping project was determined by two factors: our own assessment of interesting topics and stakeholder input. To start, we drafted a shortlist of possible prototyping topics including several provisions of the AI Act. Subsequently, we provided this list to several stakeholders of the KCDS and asked for their respective preferences. Stakeholder feedback underlined the critical role of human oversight when implementing AI systems, so we decided to organise this current prototyping project on the human oversight requirements arising from article 14 AI Act.

The EU underscores human oversight as a cornerstone for the safe and trustworthy development, deployment, and use of AI systems. This principle builds on earlier policy documents such as the *Ethics Guidelines for Trustworthy AI*, issued by the High-Level Expert Group on AI, and the European Commission's *White Paper on AI*.[6] Human oversight should ensure that AI systems remain aligned with human values, enabling intervention when systems behave unpredictably or present risks. Recognising its importance, the AI Act includes this principle as a requirement for high-risk AI systems under article 14, mandating that such systems can be effectively overseen during their use and incorporate mechanisms to enable human intervention, including through output monitoring and harmful outcome prevention.

As with our other projects, budgetary and logistical considerations shaped our approach, prioritising inclusive and practical engagement while maintaining a clear focus on the core obligations of article 14. Although we welcomed international participants at all stages of the project, we could not reimburse international travel expenses. International participants who could not travel to Belgium were invited to contribute virtually during the feedback phase. We relied on the voluntary commitment of participants and did not pay anyone for their participation. In the following section, we will outline the methodology and participant contributions in further detail.

5.2. Call for participants

In order to ensure a diverse and representative group of project participants, we combined a public call for participants alongside targeted invitations to organisations or actors that we believed would or should be interested in our project.

[6] High-Level Expert Group on AI, Ethics guidelines for trustworthy AI, 8 April 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai; European Commission, White Paper on Artificial Intelligence, 19 February 2020, https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

The public call for participants contained the following information:

- On the one hand we looked for "interested stakeholders/parties", incl. companies using or developing AI, (end) users, civil society, advisors, etc. This type of participant was expected to serve primarily as a test panel and sounding board. For instance, we aimed to offer AI providers the possibility to submit their AI application as a basis for the prototype(s) that would be developed during this project, while end users and other stakeholders could assess if the information provided by the prototypes would suffice their needs.
- On the other hand, we looked for "experts", which we considered to be individuals who have (practical) experience/expertise in facilitating human oversight measures in a technological context or in drafting compliance documents. Their primary function was to co-create and develop the prototype compliance documents. Through participation, we aimed to provide them with the opportunity to engage with interested stakeholders and improve their skills.

We expected the efforts of participants to be different depending upon whether they were a(n) (end) user/provider of high-risk AI systems or an expert. In terms of time investment, we estimated that experts would spend about 2 to 3 working days in total (attendance design workshop, further elaboration of prototypes and intervention in the feedback phase III). Other participants would probably have been able to manage with a more limited time investment, as they were not expected to contribute to the further elaboration of the prototypes. In practice, however, these roles were not strictly applied and there were several groups that collectively further elaborated their prototypes.

5.3. Phase I: Design Workshop

As a first step, we organised a legal design workshop which was conducted in-person in order to ensure meaningful personal interaction. 14 participants and 4 facilitators worked together for an entire day in three different groups to shape first versions of different prototype compliance documents. Every group focused on applying the human oversight requirements of article 14 AI Act to a single use case. These use cases were provided by providers/developers of AI-technology involved in the exercise and based on their own AI-applications (in development). This ensured that the prototyping exercise had a sufficiently concrete angle and that we could truly test during the workshop how feasible and practicable it is to integrate measures that ensure human oversight, both for the developer and the deployer of the AI system. We did not expect participants to release technical or sensitive details in relation to their use cases. It should be underlined that we looked for high-risk AI systems under the AI Act. If there was any uncertainty about whether an AI system qualified as high-risk in a specific use case, participants were instructed to assume that the use cases was high-risk for the purpose of the workshop.

The use cases are explained in-depth below (part 7). All three groups worked on prototype documents detailing the human oversight measures as if they were the provider of the high-risk AI system. These prototype compliance documents can be found in the annex to this report.

[7] Legal design experience: https://data-en-maatschappij.ai/en/event/workshop-legal-desgin

The workshop followed a legal design methodology, building further on our previous experiences with legal design workshops[7]. In practice, this means that the workshop had four parts.

1. Empathise

The first part focused on understanding the technical use case and its environment. It also included mapping the affected stakeholders for every use case (incl. users) and their concerns.

2. Define

During the second phase, participants defined the problem(s) that needed to be resolved. This included considering questions such as: what must be in the prototype? Which (legal or practical) requirements may be difficult to include? Are there aspects of the system's environment or users that are an issue for the prototype? When will the prototype be used?

3. Ideation

The ideation phase served to brainstorm about possible solutions to the problems defined in the previous phase, while taking into the affected stakeholders and their concerns into account. At the end of this phase, possible solutions were clustered, prioritised and a choice was made regarding the prototype that would be developed.

4. Prototyping

During the last phase, participants started to work on an actual prototype. As participants knew that prototypes would be further developed during the next stage in the policy prototyping project, they focused on agreeing on the structure and substantive foundation of the prototype.

5.4. Phase II: Further elaboration of prototype compliance documents

The design workshop was followed by a second phase, during which the prototype compliance documents created during the workshop were further developed by the respective team members. This phase took place during the autumn of 2024.

The idea of this phase was to create well-developed prototype compliance documents. The documents should approach, to the estimation of the participants and in so far as possible for the use cases, a final document that could have been created by a provider. The document was subsequently presented to reviewers for comprehensive feedback.

5.5. Phase III: Feedback phase

Once the prototype compliance documents were delivered, we launched phase III of the policy prototyping project: the feedback phase. In order to diversify potential feedback, we published a second call for participants. This call did not distinguish between types of participants and aimed at attracting professionals and experts in AI. People who signalled their interest to participate during the first call for participants but could not attend the design workshop were also invited to contribute to this phase. Diverse profiles responded to the feedback call (including technical or legal experts, academic researchers, an educational expert and consultants).

We gathered feedback on both the created prototype compliance documents as well as the related legal requirements from article 14 AI Act. Participants were able to provide feedback *(i)* on how the prototypes implemented the requirements of the AI Act, and *(ii)* on the practicability, feasibility, desirability and understandability of the legal requirements themselves. With regard to this second aspect, we especially tried to solicit feedback from participants who took part in earlier phases in order to capture their view on the implementation of the AI Act requirements into their own prototype. Unfortunately, only one participant to the design workshop was able to take part in the feedback phase.

Feedback was gathered through (online) interviews. A total of 11 interviews were conducted using a predetermined questionnaire which was provided to the reviewers in advance, along with the prototype compliance documents.

5.6. Phase IV: Report – publication of feedback and lessons learned

This report is the final stage of our policy prototyping project. It contains our findings, based on aggregated participant feedback, and lessons learned regarding the implementation of the human oversight requirements.

This report is driven by multiple objectives. Primarily, it aims to assist stakeholders and professionals to effectively operationalise the human oversight requirements of the AI Act by offering examples of compliance documents, coupled with best practices and valuable lessons learned. Additionally, it seeks to convey the insights gathered from this project to policymakers and authorities, providing them with a practical perspective that could be instrumental in improving the AI Act's future implementation. Lastly, the report contributes to the evolving conversation on policy prototyping, advocating for its significant value as a tool in the policy development process. Hence, **the intended audience of this report** is *(i)* policymakers involved with the AI Act and its implementation, *(ii)* supervisory authorities that will be involved in the enforcement of the AI Act, *(iii)* all stakeholders that will need to comply with, or benefit from these requirements and, *(iv)* all other interested parties.

6. THE AI ACT – TEXT OF ARTICLE 14 AI ACT

Below we include the text of article 14 as used by the participants.

§1 General Requirement
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 they are in use.

§2 Underlying goal
Human oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular where such risks persist despite the application of other requirements set out in this Section.

§3 Considerations and types of oversight
The oversight measures shall be commensurate with the risks, level of autonomy and context of use of the high-risk AI system, and shall be ensured through either one or both of the following types of measures:

(a) measures identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;
(b) measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the deployer.

§4 Functionalities
For the purpose of implementing paragraphs 1, 2 and 3, the high-risk AI system shall be provided to the deployer in such a way that natural persons to whom human oversight is assigned are enabled, as appropriate and proportionate:

(a) to properly understand the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;
(b) to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;
(c) to correctly interpret the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available;
(d) to decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system;
(e) to intervene in the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state.

§5 […][8]

In addition, **recital 73** gives guidance on how the obligation of Article 14 should be implemented. The recital repeats that the high-risk AI system should be designed and developed in such a way that natural persons can oversee their function, ensure that they are used as intended and that their impacts are addressed over the system's lifecycle. This includes an obligation for the provider to identify the human oversight measures before the system is placed on the market or put into service for the provider of a high-risk AI system. The requirement could translate into operational constraints built into the system which cannot be overridden by the system itself whereby the system is responsive to the human operator.

The natural person to whom human oversight has been assigned needs to possess the necessary competence, training and authority to carry out that role. Mechanisms that include guidance and information should be included so that the natural person can make informed decisions about if and how to intervene in order to avoid negative consequences or risks or stop the system if it does not perform as intended.

It should be noted that participants in the design workshop were provided with both the full text of article 14 AI Act as well as a simplified overview of the article. The text of recital 73 was not provided to participants, but the four facilitators present at the workshop were familiar with its content.

[8] Not applicable to the available use cases, so this was left out.

7. RESULTS

In the first phase of the project, we invited technical professionals and experts in AI and the AI Act to participate in a design (co-creation) workshop. The goal of the workshop was for participants to:

- examine and assess the human oversight requirements in the AI Act in detail;
- create operational documents, including prototype decision-making processes and prototype instructions, for human oversight.

During the workshop, participants started outlining and creating the prototype documents for human oversight. The prototype documents are based on three use cases and were each created by a different group of experts and professionals present at the workshop. The groups working on the second and third use case focused more on an overall approach to how human oversight measures could be implemented, whereas the first use case group's approach was to deliver Instructions for Use (IFU) on human oversight. This difference in approach is also reflected in the end result.[9]

The prototype documents, as provided to participants in the feedback phase, are included in the Annex to this report. We recommend having them available as a reader while reading the feedback on the specific prototype documents as feedback may refer to specific wording used in the prototype documents.

Lastly, regarding the terminology used in this report, **a distinction is made between the *deployer* and the *user* of an AI system**. The term *deployer* refers to the overarching organisation or legal entity (e.g., a company, educational institution, government body, etc.) that makes the decision to integrate and utilise an AI system within their operations. In contrast, the term *user* specifically denotes the natural person who directly interacts with and operates the AI system. This user might, for instance, be an employee of the deployer. It is important to note, however, that in some cases, the deployer and the user may be the same individual, particularly in contexts involving small-scale operations or individual proprietors.

7.1. Use Case 1 - AI education application for student feedback

The AI system in this use case aims to provide students with real-time feedback on their assignments, based on assessment criteria determined by teachers. Teachers get assessment reports that describe how the students are performing across the different criteria in a specific assignment. The assessment criteria are central to the application and act as an interface between the AI system that evaluates them and the teaching practices of the educational actors. The primary objective of the AI system is not to assign grades but to provide constructive feedback to students that improves their learning.

[9] Article 13(3)*(d)* AI Act explicitly states that IFU must address the human oversight measures outlined in article 14, including the technical measures implemented to facilitate the interpretation of the output of high-risk AI systems by the deployers.

The compliance document for the first prototype was, according to the reviewers, generally complete and well-structured, making it a good start for further integration of the human oversight obligation. Reviewers praised its clarity, user-friendliness, and detailed risk descriptions.

7.1.1. User-friendliness and informativeness

Key elements that ensured user-friendliness included:

- A **clear structure** of the document and its well-organised **table of contents**;
- A format resembling a **user manual**;
- A **logical flow**, starting with general information and gradually becoming more detailed, and;
- Comprehensive **details on risks** addressed by human oversight measures.[10]

Regarding the included **content** within the compliance document, the reviewers were generally satisfied with **the level of detail**. They highlighted several well-addressed aspects while also noting areas that could benefit from additional guidance. The table below outlines both the praised elements and those identified as lacking.

Included and found beneficial	Lacking or requiring further elaboration
• **Background information** on the purpose of the system • **Interpretative guidelines, use case descriptions and purpose descriptions** of the AI system • **Clear stop procedures**	• Details on **support and contact** channels • Clearer delineation of **what contexts** were suited for the use of the AI system (e.g., which kind of assignments for students, which age group) • Instructions and steps in the prototype, including what actions to take in case of **anomalies** (e.g., what will an anomaly look like or how should a user respond?) • **Sample scenarios** (with screenshots) and testing scenarios could be included • **Warnings** about unknown elements, such as user misuse and the impact of new data • Description of the **statistical detection of outliers and statistical monitoring tools** (e.g., how are the outliers and anomalies identified by the system)

[10] This was nuanced as reviewers explicitly stated that some risks (e.g., automation bias and over-reliance on the AI system) were not addressed adequately.

Despite identifying certain gaps, reviewers generally found the compliance document to be **fairly complete**. They acknowledged that further elaboration and additional details would be necessary as the prototype evolved and the technical implementation of the human oversight mechanisms progressed. However, they viewed the compliance document as a strong starting point. Reviewers also agreed that its clear structure enhanced user-friendliness, which was considered essential for users (i.e., teachers) to understand the associated risks and the AI system's output. This, in turn, would enable them to exercise human oversight effectively. Other key factors contributing to user comprehension included the choice of language and the use of interface design and visuals.

The **language** used in the first prototype compliance document was overall considered accessible and understandable. However, a couple of reviewers pointed out several terms used in the prototype document that might be too technical and not easy to understand for non-technical profiles. This includes use of the term "neuro-symbolic AI" and reference to the "proSLM" method, as well as other terms which can be considered expert language. Reviewers considered that this could be off-putting or pose difficulties for non-technical users. It was thus suggested to limit the use of such expert terms as much as possible, both for non-specialists in the specific sector and for non-tech-savvy users. On the other hand, one reviewer did consider that, although the language was technical, higher education instructors should be expected to understand it.

Additionally, it was suggested to add **visuals, screenshots and/or wireframes** to better illustrate the use of the system, the implementation of the measures and/or wireframes. This ties into the comment that the **interface** should be different depending on the different "users" of the AI system.

7.1.2. Risk identification & mitigation

While the compliance document addressed potential risks of using the AI system, reviewers felt some risks were not (adequately) addressed, such as **overreliance** by users (teachers) and students, or students trying to **reverse-engineer the output**. They also suggested making it explicit that the tool should not be used for grading and recommended providing more details on security measures, such as how data received by teachers is anonymised. Lastly, although the compliance document included a flagging system for students, reviewers found its description and built-in safeguards (e.g., to prevent overflagging by students) insufficiently detailed. Some also expressed a preference for a more detailed explanation of student feedback loops.

7.1.3. Proportionality and compliance

The proposed human oversight **measures were found to be proportionate or commensurate to the risks of the AI system**. Reviewers linked this to the sufficient explanations regarding the risks and misuses of the system. It should be taken into account, however, that the prototype measures have to be re-evaluated depending on the (domain-specific) context of use of the AI system and the affected person (e.g., the age of the student or grade can be determinative). Another caveat that was made is that the extent to which the prototype measures are commensurate, depends on how quickly modifications to the AI system can be made (e.g., after feedback of multiple students), as this may prove too much work for a deployer or user to accomplish within a reasonable time period. Hence, a reviewer suggested to include an estimation on how quickly changes to the system could be made in the compliance documents.

There also was debate amongst the reviewers whether the compliance document and **the proposed measures reduced the associated risks to "a desirable level"**.[11] Additional information on the implementation of the AI Act was found to be needed to definitively assess whether the oversight is as effective as required. It is also logical that the human oversight obligations, apart from the compliance documents, must still be integrated on a technical level. An interesting connection was made to the AI literacy of the individuals performing the human oversight, since if they are not adequately trained to spot anomalies, the proposed measures will prove ineffective.

As a sidenote, it is important to note that this group focused on drafting a document that could serve as Instructions for Use for users of the AI system during the design workshop. This approach appeared to enhance the user-friendliness of the compliance document by making it more concrete and actionable. The reviewers stressed that use case 1 is the right approach to informing the user about the human oversight measures.

7.2. Use Case 2 - Cardiovascular imaging

This second use case concerns an AI application that enhances ultrasound images of cardiac microvasculature systems (i.e., a network of small blood vessels in the heart that supply oxygen and nutrients to the cardiac muscle). It plays a crucial role in diagnosing heart function and disease. Medical imaging, in particular ultrasound imaging of cardiac microvasculature systems, typically renders a lower resolution image compared to the actual raw signal which contains much more data. While traditional imaging systems are not capable of rendering higher-resolution images, rapid progress is being made in AI-supported approaches. In this use case, an AI system is used to generate higher-resolution images based on the raw signal as well as detect specific anomalies within the data. In addition, the AI system could also provide an automatic diagnosis to support the cardiologist's diagnosis of the image.

This prototype document takes a different approach than the previous use case. It contains a set of oversight measures structured in a more descriptive manner. The document was still found to be user-friendly, and reviewers praised its use case description, the guidance on how to interpret the AI system's output and the tiered approach to oversight.

7.2.1. User-friendliness and informativeness

Reviewers indicated that the document as a whole is user-friendly, even with its more descriptive approach. The prototype features a clear structure, accessible and concise language, an effective visual that enhances its usability, and well-organised content. Some reviewers considered the informativeness of the prototype document to be inconsistent, particularly when compared to the first prototype, which offered more detailed background information on risks, objectives, and usage.

[11] Article 14(2) AI Act itself does not determine a specific level of risk reduction that must be achieved, therefore complicating statements concerning whether or not a prototype meets the particular sub-requirement.

The grid below outlines both the praised elements and those identified as lacking.

Included and found beneficial

- **Comprehensive and actionable** content
- Detailed **use case description** guidance
- Inclusion of **support channels**
- **Watermarking and preconditions** as support measures for human oversight
- Inclusion of a **recommended user profile**
- Inclusion of guidance on how to **interpret the AI system's output**

Lacking or requiring further elaboration

- Clarity on specific **actions** users should take (e.g., due to specific thresholds or absent preconditions)
- Details on the **capacities and limitations of the AI system**
- Details about **preconditions** (if these preconditions are not fulfilled, the system should not be used)
- Simplify the **technical jargon** and add examples that support the understanding of the system's output
- The specific **risks** and their respective mitigation measures

The elements above highlight aspects that reviewers either found beneficial or lacking entirely. Additionally, they pointed out areas that were covered in the compliance document but not in sufficient detail according to them. For instance, the reviewers suggested improving descriptions of the problems the system may cause (i.e. the description of the risks) and their potential solution through human oversight measures.

The **language** used in the context of the system is primarily medical, and this also applies to the required input, which is extensive. Some reviewers expressed concerns about the input process, as it requires a lot of time and a deep knowledge of medicine and AI. The terminology is heavily rooted in the medical field, necessitating a professional background to fully comprehend the system's output. Although the recommended user profile in the prototype was considered an added value, it also raised further questions among some reviewers. They wondered whether the user profile and **the natural person performing human oversight needed to be the same person**. For example, if the oversight is carried out by a medical professional, they might lack the technical expertise to fully understand the AI system, while a technical expert might struggle with the medical terminology. This apparent discrepancy highlighted the need for sector-specific standards or clearer guidelines on human oversight, including who should perform it and under what circumstances. Additionally, the skill set of the users themselves was called into question, particularly regarding how to restrict system use to appropriately qualified individuals given the high-stakes medical context.

Some gaps in clarity were noted, particularly concerning user control over the data considered by the system and how actual oversight would be ensured. Concerning the latter, a reviewer suggested external audits or supervision of those overseeing the functioning of the system.

7.2.2. Risk identification & mitigation

A key strength of the prototype are the **comprehensive risk mitigation measures**. Reviewers indicated that the prototype allows to identify risks (although, as mentioned above, this may have been explored in more detail for some reviewers) and provides actionable solutions. Risk mitigation steps such as a reporting system and feedback channels were appreciated in this regard.

Reviewers offered diverging views on the prototype's description of its **capacities and limitations**. While some found the explanations sufficient, others felt they lacked detail. Clearer descriptions were suggested to help users better understand the system's potential and constraints. The reliance on assumed user expertise was flagged as a limitation.

While the description of how to **interpret the AI system's output** was seen as positive by some participants, there was also criticism regarding the explanation of how to correctly interpret output. One reviewer noted that this aspect was missing in the prototype, while another warned that too much reliance is placed on training and suggested that the measures for interpretation should be distributed more evenly.

Automation bias emerged as a shared concern. Although the prototype acknowledges this risk, reviewers called for more robust measures to address it. Some reviewers felt that the suggestion of a tiered approach, where a second person is involved in the oversight process, is sufficient and well-described. Others criticized the lack of specific attention to automation bias and requested more input on measures such as rectification procedures.

Concerns about **potential misuse or overreliance** on the system surfaced. While robust when used as intended, the prototype's reliance on user expertise was considered to raise the risk of **under- and overutilisation or inappropriate delegation of tasks**. Additionally, the prototype's design places significant demands on cardiologists, further contributing to concerns about underuse or misuse of the system. Adding measures to address these concerns would be a valuable addition to the prototype.

7.2.3. Proportionality and compliance

The **proportionality of the oversight measures** to the risks and autonomy levels associated with the high-risk AI system was evaluated positively by most reviewers. It is emphasised, however, that this may also depend on the context in which the system is implemented and the involvement of developers in the process.

The prototype's human oversight measures were generally regarded as effective in mitigating risks to health, safety, and fundamental rights. The tiered approach is seen as good practice. Two reviewers noted that it might be worth considering involving the additional person earlier in the review system, although this might make the decision-making process more complex. As it would add to the workload, there is a chance that cardiologists would drop out and reject the use of the system.

Overall, a majority of reviewers found the second prototype compliant with the AI Act requirements. A minority of reviewers doubted that the documents and measures would comply with the AI Act, saying that the documents and measures were not sufficiently detailed and that additional explanations were required.

7.3. Use Case 3 – Big data policing

Big data policing is an innovative strategy that uses historical data to forecast when and where there is a high risk of new crime events (residential burglaries) in order to use police resources more efficiently and proactively and ultimately reduce crime rates.[12] This use case centers around human oversight measures for a big data policing model which recommends patrol routes. Big data policing models can consist of variables based on crime data available in police databases (e.g. previous crime events), socio-economic data (e.g. poverty index, residential mobility), opportunity characteristics (e.g. the presence of shops, distance to the nearest highway), data from new technologies (e.g. intelligent cameras) and other known predictors of crime (e.g. police patrol intensity).

This third prototype did not have a clearly defined strength according to reviewers but was instead generally considered insufficiently detailed and its measures left room for improvement. The prototype differs from previous ones, in terms of format and structure, and focuses specifically on training and an infographic to enable human oversight. While several respondents commended these measures as a good start, they also identified gaps in usability, risk identification and mitigation, and compliance with the AI Act which the prototype should have addressed.

7.3.1. User-friendliness and informativeness

This prototype was considered less **user-friendly** than use cases 1 and 2. Particularly the document's lacking structure, (expert or high-level) language and the description of the system's risks and purpose were considered insufficiently user-friendly and complete. The lack of sufficient description of some of those elements was also found to hinder the prototype's **informativeness and effective oversight**. The grid below outlines both the praised elements and those identified as lacking.

The identified gaps in informativeness and effective oversight led reviewers to conclude that the prototype, while a decent starting point, still needed to address various issues to improve its compliance with the AI Act. With regard to the organisation of the oversight, several reviewers highlighted concerns about how the oversight would fit into existing organisational workflows, and how and when the system would be deployed. Additionally, while reviewers recognised that the measures could improve accountability, they may also **slow down processes or impose additional burdens on staff**. One challenge for the prototype is to integrate oversight measures that are sufficiently rigorous yet do not erode potential efficiency gains.

Two oversight measures in use case 3 received specific comments from reviewers, namely the infographic and the proposed training for users. The concept of the **infographic** was considered user-friendly by reviewers to allow understanding of the AI system although its usability as a human oversight measure was questioned.

[12] In the policy prototyping project, this use case was considered high-risk under art. 6 (2) and Annex III, 6(a) AI Act

Included and found beneficial	Lacking or requiring further elaboration
• **Inclusion of KPI's** related to the system • **Different forms** of possible oversight (collective and individual) • The **feedback "by confirmation"** mechanism for patrol officers	• Description of **risks** (e.g. fundamental rights, profiling, discrimination loops enforcing biases,...) • Information on potential **malfunctions** of the system and possible corrections by overseers • Description of **data** used in the system • Description of the **governance structure** of the oversight (e.g. interactions between patrol officers, dispatchers, etc.) • Details to allow understanding of the **capabilities of the system** • Insufficient measures to address **over-reliance on the system**

While some reviewers found that it gave insights into the system's operational logic such insights were not considered to help overseers identify errors in the system or recognise automation bias or misconceptions. Additionally, whether the information provided was sufficient for a user to determine why a route was suggested (and subsequently oversee this decision) was not clear for reviewers. Finally, reviewers suggested incorporating elements from the training into the infographic to assist the oversight.

Reviewers' opinion on the training component was mixed. Some reviewers considered the **training component** described in the prototype a major strength that can equip officers and other potential users with the skills needed to interpret and monitor the system's outputs. The training could help mitigate automation bias and overreliance on AI recommendations. However, several other reviewers emphasised that training alone, while crucial, can be an imperfect tool and that the oversight should not, for its majority or solely, rely on the training course. To be effective, these reviewers stressed that the specifics of the training needed to be appropriate for the different roles and should include concrete examples of inaccurate outputs of the systems and ways to resolve them.

7.3.2. Risk identification & mitigation

Reviewers suggested developing a clearer risk inventory and more clearly assigning proportional specific risk mitigation measures to better oversee or resolve specific risks. Such measures should also clearly describe the required actions by the patrol officers, e.g. with regards to acceptable inputs and dysfunctions of the AI system. The provider should ensure that the training materials provided to users are sufficiently representative and relevant to well-defined users. In relation to overreliance on the AI system/automation bias, reviewers found that officers may struggle to question AI recommendations in real-world policing scenarios without clear instructions or easily accessible decision-support tools, despite the existing "by confirmation" in the prototype.

One reviewer also proposed incorporating **citizen participation** or **external oversight** mechanisms as measures to mitigate foreseeable misuse. Community oversight panels or similar bodies may bolster public trust and ensure that operators remain accountable when making decisions with high stakes for individuals and communities.

7.3.3. Proportionality and compliance

This third prototype met the requirements of article 14 AI Act the least compared to the other two use cases. It provided significantly less information, making it more difficult for users to understand the AI system's functionality, limitations, and potential risks. The lack of clarity in its documentation may hinder effective human oversight. Additionally, several measures were found to be lacking or insufficient to proportionally address the risks posed by the AI system. Further refinement and consideration of the measures and the document are needed to address the risks.

Interestingly, **pilot testing** emerged as a recurring suggestion. By experimenting with the prototype in controlled environments, developers could identify training pitfalls, data inconsistencies, or user difficulties that might otherwise go unnoticed. This small-scale testing could, for instance, reveal how quickly officers become reliant on the AI's suggestions, or how well they can detect errors in real-time.

7.4. General comments by reviewers on the prototype documents

The significant difference between the three use cases proved highly beneficial, as each angle offered distinct and valuable insights. The reviewers universally agreed, however, that the prototype document from the first use case was the most clear, informative, and complete. Interestingly, the drafting approach taken by this work group differed slightly from the others. They prioritised creating instructions for the deployer and user of the AI system rather than focusing solely on fulfilling the general human oversight obligation. This emphasis on providing clear and actionable instructions seemed to enhance the user-friendliness of the prototype documents.

The structure of the documents varied across the use cases, reflecting the **uniqueness of each sector**. This tailoring was viewed positively, as it ensured sector-specific relevance. For example, the third use case which involves law enforcement professionals, accounted for the likelihood that these users might have limited knowledge about AI systems. Nonetheless, it must be acknowledged that these preconceptions are not necessarily accurate. Reviewers remarked that it seemed that users from law enforcement were considered to be less advanced in using AI than the cardiologists from use case 2 for example. The true risk lies therefore in overestimating users' AI proficiency, which could result in ineffective human oversight measures. Contextual differences mean that appointing individuals for oversight is not a one-size-fits-all solution. Each organisational ecosystem requires a tailored approach, factoring in all relevant AI Act obligations.

The prototype documents were generally deemed sufficiently informative as a starting point, though reviewers emphasised the **need for further clarifications and specifications**. For example, more details on the **roles and responsibilities** of the actors involved in the human oversight could be added, as well as further details regarding the actions specific actors need to undertake if the AI system makes a mistake or shows a certain behaviour.

This governance process by the deployer could (or should) even include **a second tier of human oversight**. This second tier differentiates between the type of oversight that needs to be done by the users (e.g., teachers in UC1) and the type of oversight that needs to be done by deployers (e.g., the educational institution in UC1) or by a different coordinating level (e.g., management of a group of schools in UC1). Additionally, such a second tier could include a review of the oversight performed by the user or a second instance of oversight on the AI system. In terms of suggested improvements, those reviewers also indicated that providing more technical guidance and **establishing KPIs** for the system would improve the oversight.

Another recurring concern of reviewers was about **AI literacy** of deployers and its relation to human oversight. Many deployers lack in-house AI expertise, and while AI literacy is required, the level of literacy and the associated resource expectations need clarification. This also raises the question regarding the level of AI literacy to be expected from users of the prototypes and if they would need additional AI literacy training, in conjunction with the measures of the compliance documents. This points to a **potential underlying lack of resources on the deployer's side**. This issue is not unique to one sector but reflects a broader challenge: it remains unclear what concretely is expected of deployers in fulfilling the human oversight obligation, while article 26 (2) AI Act does set out obligations for the deployer to assign human oversight to natural persons with the necessary competence, training and authority, as well as the necessary support.[13] However, this obligation does not divide the specifics of the human oversight obligation between the provider and deployer, causing confusions particularly when articles 14 and 26 (2) AI Act interplay.

The human oversight obligation must also be considered alongside article 13 AI Act, which addresses Instructions for Use (IFU). These IFUs are likely to provide additional practical guidance on human oversight, enhancing implementability for the deployer. It is clear that the human oversight obligation does not function in isolation but interacts with other obligations under the AI Act, all of which must work together for effective compliance.
In conclusion, the prototype documents represent a useful first step but will require further refinement and integration to become fully operational compliance tools.

A majority of reviewers found that the prototype compliance documents included all the requirements listed in Article 14(4) AI Act and were positive about some of the included measures. Reviewers did, however, also note several points of improvement. For instance, the measures described in the prototypes were still considered **vague** on many aspects. Other reviewers indicated that the threshold to determine if a measure is sufficiently appropriate under the AI Act is difficult to pinpoint and cautioned against making the measures **excessively burdensome** on the user, particularly to reduce the chance of automation bias. Another recommendation by reviewers to improve the measures was to **properly combine technical features for human oversight in the system with a description in the compliance documents to illustrate how the system should function and what the desired system behaviour should be**.

Finally, many reviewers noted that legal compliance of the prototype depends on upcoming interpretation of the AI Act by authorities and that the measures in the document need to be appropriately implemented by the deployer.

[13] Art. 26 (2) AI Act

7.5. Recommendations and lessons learned for human oversight measures and documents

Based on the extensive discussion of reviewers' feedback above, this section of the report bundles recurring recommendations, best practices and lessons learned from the policy prototyping process which can be used by providers and deployers to enhance their own human oversight measures and related documents.

7.5.1. Recommendations for human oversight

OVERSIGHT GOVERNANCE BY THE DEPLOYER

The measures and documents created by the provider should take into account the common **structure of the deployer's organisational practices** and the different users likely to interact with the system. To guide the deployer's oversight of the AI system, the compliance documents should build on the different roles in the deployer's organisation and the different tasks and responsibilities they should take up in the oversight. For example, in a school or hospital, teachers or doctors may be expected to perform day-to-day oversight while the school's or hospital's director or managing board, in cooperation with the provider, may define the guidelines determining concrete human oversight measures. This description of possible governance at the deployer's side can include measures that enable a second person to perform oversight on the results of the AI system or could include an additional "appeal" step which allows affected persons to contest the decision made by the person responsible for the oversight (e.g. a student/ patient who believes that a teacher/doctor should have intervened in the AI system's functioning).

DESCRIPTION OF RISKS AND USER PROFILE

While not strictly required under article 14 AI Act, providers can improve the usability of the oversight measures and the documents by providing **background information on the AI system, its intended use and purpose and the risks** that the provider intends to limit through the human oversight measures.[14] A correct and complete description of the various risks posed by the AI system was seen as a crucial element in understanding the oversight measures and the associated actions for the users and deployer. Similarly, a description of the envisaged or required user of the AI system (and specifically their qualification or competences) was considered important for the correct functioning and oversight of the AI system.

[14] To the degree this information is not yet included in IFUs drafted under article 13 AI Act.

MEASURES AT TECHNICAL LEVEL

Providers should **integrate the human oversight measures** into the AI system, its functioning and user interface as much as possible. These types of included measures were considered more likely to be effective than organisational measures, in addition to being easier for deployers.
These measures may include, for example, reminders, nudges and checks by the system to verify that oversight is being performed as well as automated means and procedures to guide the oversight and guidance included in the system on what actions the deployer should undertake.

AVOIDING AUTOMATION BIAS

In terms of specific human oversight measures, the provider should take special care to include adequate measures to **avoid automation bias and overreliance** on the system. This was a recurring concern expressed by reviewers for all use cases. It is important for users to grasp the functioning of the AI system to not feel overwhelmed and default to the system's recommendations. Additionally, the inclusion of a second overseer and approaches in which the human has to make a first decision on the matter (e.g., a diagnosis of a patient) before being provided with the output or recommendation of the AI-system were found to be beneficial.

TRAINING

Training of users and deployers was seen as a very useful way of supporting human oversight. Such trainings need to be appropriate for the different roles and should include concrete examples of inaccurate outputs of the systems used and ways to resolve them. At the same time, it is not a guarantee for the effective exercise of oversight measures on its own.

STOP BUTTON

Finally, providers may include **a "stop-button" to halt the functioning of the AI system** if appropriate for the system. This was less relevant in the use cases considered during this project, which focused less on AI systems where such a stop button would result in an immediate mitigation of risks. However, both providers and deployers may also consider a "stop-button" at organisational level, indicating circumstances in which the deployer should stop the use of the AI system for example because the system is no longer suited for their purposes or because the outputs have become unreliable.

7.5.2. General best practices

BALANCE BETWEEN DETAIL AND INFORMATION

Providers have to balance the **amount of information and detail** they provide with the deployer’s capability to understand and process the information. Providers should include both high-level as well as detailed descriptions of how the human oversight measures (are intended to) function and should be implemented. For example, providers may want to describe the limitations of the AI system both in technical terms (statistical explanations) as well as in layman's terms ("the AI system is not suited for use on people older than 40"). Where necessary, the documents should very practically describe the actions that a user should undertake to optimally perform the oversight (e.g., intervening at a certain value, processing or inputting information in a certain way, recognizing a specific mistake of the AI system etc.).

LANGUAGE AND ADDITIONAL INFORMATION

In order to guarantee the usability of the oversight measures, the language used should be **adapted to the target audience(s)**. This may include providing separate sections for different users depending on their technical knowledge and the role they fulfill in the oversight.

STRUCTURE

The structure of the compliance documents was deemed to be of great importance. A **clear and logical structure** increased the user-friendliness of the documentation and was considered useful for deployers. This includes structuring documents with **a table of contents** and maintaining **a logical flow** from general to more detailed information on the AI system and the oversight measures. In addition to a clear structure, reviewers also found that **visuals, wireframes and screenshots** could improve the readability of the documentation and the understandability and usability of the measures for deployers.

8. FEEDBACK ON THE AI ACT

In addition to providing feedback on the prototypes, reviewers were also asked to provide feedback on article 14 AI Act. They could provide feedback on the practicability, feasibility, desirability and understandability of the article. These terms were used as follows:

Practicability:

assesses whether the requirements can be implemented without excessive difficulty in real-life situations.

Feasibility:

assesses whether the requirements can be operationalised given the resources and constraints available to a provider, such as budget, time, technology and manpower.

Desirability:

assesses whether the requirements and their operationalisation are useful and valuable to the intended audience and users.

Understandability:

assesses how well the proposed requirements can be understood by the intended audience.

8.1. Practicability

Only a minority of reviewers considered that the requirements were sufficiently practicable, realistic and provided enough leeway to providers and deployers. A larger group of reviewers considered that the practicability of the article could be improved in various ways. Notably, the most recurring comment by reviewers was about the **unclear and broad meaning** of the provision. Some reviewers believe that, with expected additional guidance and examples, the text of the article is not overly difficult to comprehend. They recognised that the nature of legislation often results in obligations being somewhat opaque initially. The article does provide a clear starting point, but practical implementation will require creativity and interpretation by those in the field. However, not all reviewers shared this view, with some **criticising the vagueness of certain terms** in the provision, such as '*commensurate*', '*levels of autonomy*', '*reasonable foreseeable misuses*', and '*context of use*'. The proportionality and feasibility of human oversight measures were noted as adding further ambiguity to the obligation. This room for subjectivity was viewed by many as problematic, as it creates the risk that some providers or deployers may take the path of least resistance to comply with the requirements rather than ensuring robust processes and instructions. Ultimately, room for interpretation can lead to misinterpretation, which could undermine the effectiveness of the oversight. According to one reviewer, the article establishes principles rather than specific, actionable obligations for actors.

This was also reflected in the feedback on the various prototypes. Reviewers generally agreed that the prototypes complied with the AI Act requirements, while at the same time providing many points of feedback to improve the compliance. This suggests that the practical implementation of the human oversight can be challenging and highlights the uncertainty that arises from broadly worded provisions such as article 14 AI Act, specifically around the "minimum" level of compliance. Despite this criticism, the reviewers also provided positive remarks such as stating that the requirements were a good start or that companies are provided with a decent amount of freedom in determining the oversight measures.

Reviewers indicated that (legal) certainty and consistent implementation of the requirements was important. To ensure this, guidelines should be published, as the alternative would be to wait for case law. In this regard, the authors of this report point out that a related standard is currently being developed by CEN/CENELEC. The lack of clear benchmarks and standards was also criticised by these reviewers. They considered this a significant shortcoming, as it makes it harder to establish consistent practices across sectors. It is unsurprising that the AI Act addresses this obligation only at a general level, as its practical implementation will need to be tailored to specific sectors and contexts. Translating these obligations into detailed, actionable practices remains a key challenge. In this regard it should be noted that reviewer both valued the flexibility offered by less specific obligations for providers as well as felt a need for legal certainty and consistency. These seemingly opposite considerations will need to be appropriately balanced.

Reviewers generally suggested that the article, or associated guidelines, should further **specify when the human oversight requirements can be considered fulfilled**. The requirements should be made more detailed and/or actionable according to these reviewers. For example, by policy makers providing (timely) guidance such as setting out specific scenarios in which human oversight may apply, establishing (industry) standards or including specific, solid levels of required oversight. A reviewer also noted that there was a need for examples to illustrate the way forward for providers (improving also the feasibility of the requirements) and that internal KPIs indicating when an organisation complies with human oversight obligations would make compliance easier to determine.

Another recurring comment by reviewers is that article 14 AI Act imposes obligations on the providers with regard to human oversight measures but does not sufficiently assign responsibility (or set out measures) for deployers (and other stakeholders) of the AI system. Even taking into account article 26 AI Act, the reviewers found that the AI Act remains vague on which involved party (deployer or provider) should implement certain measures or determine the governance around the use of high-risk AI systems. Consequently, **the relation between the deployer's obligations and the obligations of the provider to facilitate human oversight measures is considered unclear and would have benefited from additional details**. Reviewers expressed that they wish the AI Act determines the responsibilities related to the governance of AI systems for both the deployer and provider clearer. One reviewer added that every user should retain full responsibility for their tasks, without delegating them to others (e.g., in the case of a doctor making an assessment), since their expertise was required for proper compliance with the requirements. Additionally, one reviewer remarked that it was also difficult for providers under these requirements to assess if the measures should target other stakeholders of the AI system (e.g., affected persons).

In conclusion, a reviewer highlighted the challenge providers face in assessing all potential risks and possible malfunctions of an AI system in advance, as well as identifying appropriate human oversight measures. Providers must consider a wide range of possible contexts, especially when the AI system operates across different sectors. However, accounting for all possible uses of the AI system and tailoring human oversight measures to each specific context may prove challenging, which impacts the practicability of the human oversight requirement.

8.2. Feasibility

Reviewers generally considered that the requirements were feasible for AI system providers. One reviewer specified that compliance with the requirements is expected to be feasible to a certain extent, at a minimum encompassing the identification and mitigation of the most significant risks and foreseeable misuses of the AI system by providers.

Several reviewers were concerned with a **possible lack of expertise or knowledge of the AI system by the user**. Interestingly, reviewers had varying expectations for different types of end users. For instance, a cardiologist was expected to have a greater understanding of the technical aspects of the AI compared to law enforcement officers, reflecting differing knowledge and expertise typically associated with these professions. The reviewers stressed that proper AI literacy training is a prerequisite for supporting the feasibility of the requirements of article 14 AI Act.

Reviewers consequently considered the feasibility less manageable for **startups** compared to larger companies (although one reviewer did consider the distribution of roles in the oversight process to be feasible for all deployers). A reviewer considered that providers or deployers in general may face **time or budget constraints** which prevent human oversight and that increased bureaucracy or box-ticking may also harm human oversight.

Not enough companies were considered to be sufficiently aware of the human oversight requirements, nor do they provide sufficient consideration to the requirements at the moment, which may hamper proper implementation. **Creating awareness would require resources and effort**. The reviewer suggested that providers should in some way be supported in creating this awareness, at the EU or member state level.

To conclude, a small minority of reviewers did not find that the human oversight requirements were entirely feasible. One of the primary reasons identified by these reviewers ties into the practicability of the human oversight obligation; at the moment the requirements are still **too theoretical and abstract** to assess and implement to the full extent. The reviewers considered that more time and clear standards would increase the feasibility of article 14 AI Act.

8.3. Desirability

Human oversight is widely regarded by the reviewers as a crucial element in the deployment of high-risk AI systems, particularly given the persistent challenges that arise during their use.

Reviewers mentioned that human oversight is essential to build **trust**. By incorporating human oversight, users and deployers are empowered to intervene when necessary, creating greater confidence compared to systems that do not have oversight functionalities or measures. A potential trade-off between implementing human oversight and maintaining operational efficiency was mentioned though, noting that the added oversight could impact the efficiency of AI system deployment. Balancing the oversight measures (and associated human verification) and the operational efficiency of the AI system may be difficult and this should be taken into account according to a reviewer.

Human oversight also enhances **the understanding of the working of AI systems for the parties involved**, particularly regarding their intended functionality and the circumstances under which a user might be held responsible. Article 14 AI Act is seen as an effective measure for enabling such increased clarity.

Overall, the **obligation is widely regarded as valuable**. According to the reviewers, it contributes to the more effective and responsible use of AI. While the practical added value of the obligation is less criticised, it is evident from related discussions that the obligation itself is not always entirely clear. Much will depend on further concrete elaboration and the adoption of best practices to ensure its successful implementation.

8.4. Understandability/clarity

The translation of legal requirements into technical measures, and into understandable wording for deployers, is not sufficiently addressed in the AI Act. One reviewer noted the difficulty in finding corresponding technical solutions and translating measures to the language of deployers. This underscores a common challenge regarding the language and terminology used in the context of AI systems and human oversight. A technical expert may not fully understand sector-specific terms, while a professional user, familiar with the sector, might struggle with the technical jargon involved in using and supervising an AI system. Many reviewers suggested that **a simplified translation of the technical and legal requirements** would be beneficial, especially for audiences unfamiliar with AI systems or the specific sector in which AI is deployed. However, it was largely agreed that *experts* in relevant fields find the concept of human oversight and the associated requirement in the AI Act understandable, although the aforementioned room for interpretation remains a concern. One reviewer also noted that what is not currently common knowledge might become so in the future as more users gain experience with AI systems.

8.5. Recommendations

As repeatedly emphasised, **additional information and guidance** are necessary to support the implementation of the AI Act. It was reiterated that concrete examples are essential to clarify the required measures for human oversight. Other elements that these guidelines could include, are:

- how to address frequently occurring risks and misbehaviours associated with AI systems
- best practices and lessons learned, both for providers and deployers
- an illustrative lifecycle timeline for implementing specific measures
- whether the oversight must be performed on the level of individual decisions or if it can also happen in an collective manner
- the alignment with obligations from other existing legal frameworks (e.g., General Data Protection Regulation, Digital Services Act, etc.) as ensuring alignment might strengthen the consistency and coherence of the respective (oversight) processes

Certain sectors, such as law enforcement, will require tailored guidelines, which some reviewers believe should remain non-binding to ensure flexibility based on the expertise of the involved parties. These sectors should also have the opportunity to provide input to ensure the guidance reflects their specific needs. These guidelines should be adapted as technologies evolve in order to remain effective and relevant. Open-source tools for documentation or risk frameworks to facilitate compliance and risk management would also be beneficial. Lastly, a platform for stakeholders to exchange knowledge and experiences and questionnaires that help categorise and describe AI systems to generate actionable and custom oversight recommendations were also suggested.

Equally important is fostering multidisciplinary collaboration among legal experts, technical professionals, and other stakeholders to effectively translate legal requirements into technical and actionable processes.

The importance of **standards and standard operating procedures** was widely supported. Some reviewers also expressed a need for these to align with other obligations, such as conducting a Fundamental Rights Impact Assessment (FRIA) or a Data Protection Impact Assessment (DPIA). Additionally, **regulatory sandboxes** could be leveraged to explore and refine methods for implementing human oversight in real-world scenarios.

8.6. Feedback on policy prototyping

The reviewers regarded policy prototyping as a valuable approach for exploring topics that have not been extensively studied before. By reframing obligations in a more practical and actionable way, this exercise becomes an essential tool for supporting AI deployers. However, its reach could still be broadened. One reviewer suggested including a wider range of stakeholders, such as affected individuals and students, to ensure the exercise captures diverse perspectives and needs. A reviewer described the exercise beneficial for the ecosystem, saying it could pave the way for a new approach to research on these topics that hasn't been explored before.

9. CONCLUSION

Article 14 AI Act emerges as a critical component in effectively controlling the risks for health, safety and fundamental rights posed by high-risk AI systems. This obligation aims to mitigate harmful outcomes by ensuring that humans remain in control and capable of intervening when necessary, fostering accountability and safety in AI systems.

This report provides comprehensive insights, suggestions and practical guidance for policymakers, supervisory authorities, stakeholders, and professionals who navigate the complexities of the AI Act's human oversight requirement. Through the development of policy prototypes, the report sheds light on both sector-specific and general insights into the concrete implementation of article 14 and the difficulties that providers and deployers may face. These prototypes act as practical examples to explore how the obligation can be applied effectively in diverse contexts, taking into account the unique needs and challenges of each sector. Subsequently, the report provides detailed comments and feedback on article 14 itself, aiming to inform and guide policymakers and authorities on difficulties as evaluated by the reviewers.

Below, we present the key findings of the report, based on a summary of the feedback received.

Article 14 is written in broad terms, providing flexibility for both deployers and providers. While this flexibility has its benefits for providers, for example by allowing scenario-specific measures, it also makes the practical implementation of this obligation challenging and leads to legal uncertainty. This underscores the **pressing need for (sector-specific) guidelines by authorities**, as the human oversight obligation can be fulfilled in various ways but requires a yardstick by which providers and deployers can measure their compliance with the obligation. This can be provided in the form of concrete examples, sector-specific advice or technical standards to meet the requirements. The requirements were generally considered desirable by project participants.

As a starting point, it may be helpful for providers to enable human oversight by providing **specific instructions to the end-user** situated at the deployer. This approach was (successfully) applied in the first use case, which was widely regarded as the most comprehensive compliance documentation. Additionally, it is clear that the human oversight obligation must align with article 13 AI Act, which outlines the transparency requirements and the drafting of instructions for use by the provider for the deployer.

Regarding human oversight, the **profile(s) of the individual(s)** performing the oversight is particularly important. These persons will likely need a combination both sector-specific knowledge to assess the AI system's output and technical expertise to understand (at least broadly) how the AI system operates and when it fails in its task. Given this specific profile, careful consideration must be given to who carries out human oversight, how the oversight governance of the AI system is best structured at the deployer's end, whether appeals against oversight decisions are possible, and what training the individual(s) must undergo.

This ties in with **the distribution of tasks** related to the human oversight obligation among the provider and deployer. This requires establishing an effective governance structure that ensures alignment between the provider and deployer, allowing them to complement each other's roles.

Providers should **designate specific roles in the deployer's organisation** who are particularly suited to perform certain oversight actions. This does require providers to investigate common organisational roles and structures at the deployers they provide systems to. Closely linked to this consideration is the **choice of language** used in the instructions provided to the individual performing human oversight. This language will need to include both technical and sector-specific terminology focused on the specific role that handles the instructions.

However, the text of the article does not clearly define this distribution of tasks, which has been noted as one of the more challenging aspects of implementing the human oversight obligation. **Additional guidance** on the content of the human oversight obligations and the distribution of responsibilities among the parties by authorities could be a significant help in this regard and provide clarification on which actor must perform which tasks.

Lastly, providers could follow the other **measure-related best practices** which reviewers in this project have found useful for example relating to addressing **automation bias or providing training**. General best practices, such as providing additional information and extensive background on the AI system and its functioning or ensuring a logical and clear structure of the compliance documents can also benefit providers and deployers. Authorities and policy makers may additionally consider measures to increase awareness among providers and deployers of the requirements and the need to work on compliance, as well as measures to increase AI literacy at those parties to increase the feasibility of the requirements.

10. ACKNOWLEDGEMENTS

We would like to thank all participants whose contributions made this policy prototyping project possible. Your enthusiastic engagement and valuable insights were pivotal in shaping this project. Special thanks are extended to those who participated in the design workshop, investing time and expertise to draft mock documents. The collaborative efforts of everyone involved have been instrumental in the production of this report.
Your commitment to advancing the discourse on policy prototyping in the field of AI and data policy is genuinely appreciated.

Project Participants

Rafa Galvez
Postdoctoral researcher cybersecurity and AI – COSIC, KU Leuven

Anastasia Karagianni
Doctoral student - LSTS VUB, FARI

Nathan Genicot
Researcher - LSTS VUB

Fred Lefever
EdTech applications – ICTS, KU Leuven

Victor Frenay
Experience researcher - IO

Martin Canter
AI and Data Expert - FARI

Eva van Bree
Compliance specialist - IMEC

Pieter Gryffroy
Attorney-at-law, Timelex

Luca Nanini
AI Governance Consultant - Minsait

Helen Tueni
Founder - Vucable

Cedric de Koker
Researcher AI and law - VIVES

Bart Magnus
Expertise officer - Meemoo

Jonas Haspeslagh
Business Analyst – ICTS, KU Leuven

Alen Katrien
Kenniscentrum Digisprong

Mihnea Vlad Turcanu
Research Associate - Medical Imaging Research Centre, KU Leuven

Ingrid Lambrecht
Expert advisor - Legile

Dries de Roeck
Product Manager - Helpper

Dafna Burema
Postdoctoral Researcher - TU Berlin

Naomi Theinert
Doctoral Researcher, UGent

Pavlos Sermpezis
Postdoctoral Researcher - Datalab, Aristotle University Greece

Maaike E. Harmsen
PhD researcher, Ethics and new technology - U Amsterdam

Laura Lucaj
AI Auditing and Regulation Researcher - VW AG

Hans Arents
Senior advisor digital government - Flemish Government

Bilgesu Sumer
Doctoral Researcher - Centre for IT and IP law

Knowledge Centre Data & Society Team

Wannes Ooms
Researcher – Centre for IT and IP law - KULeuven/KCDS

Lotte Cools
Researcher – Centre for IT and IP law - KULeuven/KCDS

Thomas Gils
Research Associate – Centre for IT and IP law - KULeuven /KCDS

Frederic Heymans
Research Associate – imec-SMIT, VUB/KCDS

Knowledge Centre
Data & Society

Artificial
Intelligence
Flanders