AI Office's Consultation - Future-Proof AI Act: Trustworthy General-Purpose AI

Comments and Remarks by Selbstregulierung Informationswirtschaft e.V.

September 2024

selbstregulierung informationswirtschaft e.V.

Publisher

Selbstregulierung Informationswirtschaft e.V.
Großbeerenstraße 88
10963 Berlin
https://sriw.de

+49 (0)30 30878099-0
info@sriw.de

Amtsgericht Berlin Charlottenburg
Registernummer: VR 30983 B
USt-Nummer: DE301407624
Deutsche Bank AG
IBAN: DE33 1007 0000 0550 0590 00

Vorstandsvorsitz
Dr. Oliver Draf

Geschäftsführer
Frank Ingenrieth

Table Of Contents

1 Web-Form questions

1.1 Risk Assessment Related Questions

1.1.1 Question 10. List of Systemic Risks

We suggest that methodologies and risk assessment practices familiar from areas such as GDPR, IT security, and business continuity planning will be considered when evaluating AI-related risks. Integrating these established methodologies will help ensure compatibility between various mandatory assessments, thereby minimizing unnecessary administrative burdens for businesses.

We believe it is crucial to maintain a focused approach, prioritizing elements and risks that have already been identified as relevant through prior assessments and proven methodologies. While a certain degree of attention to probabilities and foreseen risks is valuable and aligns with best practices in risk assessment, the primary focus should remain on current, tangible risks.

In instances where potential risks do not yet reach a threshold of likelihood or realistic scenario, we recommend backlogging these issues to maintain a balanced approach. This strategy would help reduce unnecessary burdens on stakeholders while ensuring that attention is concentrated on immediate and significant areas of concern.

1.1.2 Question 12 Risk Assessment Measures Reflect Differences in Size and Capacity

In line with the remarks for 10 (1.1.1), we propose that methodologies from GDPR, IT security, and business continuity could serve as effective blueprints for risk assessment approaches in the context of AI regulation. These established frameworks provide valuable insights that can be adapted to ensure a robust and compatible assessment process. In this line of action adopting or adapting existing benchmarks of the market for generative models into risk assessment structures could facilitate the objective.

We appreciate that size and capabilities are recognized as differentiators in risk assessments. It is important, however, that these differentiators are closely examined in terms of how they impact the probability of risks occurring. Emphasizing probabilities and focusing on existing, proven challenges allows for a more targeted approach, directing resources toward areas of highest relevance and immediate need.

1.1.3 Question 13 Current State of the Art (Risk Assessment Measures)

Measures implemented within the AI regulation should prioritize effectiveness and actual necessity, ensuring that interventions are proportionate to the risks they aim to address. The probability of risks

occurring should play a critical role in determining the need for specific measures, helping to balance precaution with practicality.

It is important to recognize that preventive measures, while valuable, may not always be the most effective approach. In some cases, a robust reporting mechanism combined with the ability for providers to respond quickly and adaptively can offer a more efficient solution. Additionally, automatic analysis of use cases and monitoring deviations between expected and actual outputs of models may provide more precise and dynamic risk management than traditional preventive measures, allowing for timely adjustments that better reflect real-world conditions.

1.1.4 Question 15 Current State of the Art Model Evaluations

In reference to our reply to Question 13 (1.1.3), we believe that regulatory measures should be guided primarily by their effectiveness and actual necessity, ensuring that resources are directed toward genuinely impactful actions. The assessment of risk probabilities should be a key factor in determining the need for specific measures, allowing for a more tailored approach that addresses realistic concerns.

It is also important to consider that preventive measures, while often a key component of risk management, may not always be the most effective strategy. In many cases, a well-designed reporting mechanism coupled with the ability for providers to respond quickly can offer a more flexible and immediate way to manage emerging risks. Furthermore, automatic analysis of use cases and monitoring deviations between expected and actual model performance may provide more precise insights, enabling a dynamic response that preventive measures alone might not achieve. This approach supports a balanced and adaptive regulatory environment that focuses on real-world effectiveness.

1.1.5 Question 17 Greatest Challenges in Risk Assessments

In conclusion, it is essential to avoid the creation of yet another standalone risk assessment that runs parallel to existing frameworks such as GDPR, IT security, or business continuity. The introduction of additional, distinct assessments risks creating incompatibilities between different methods and the measures they require, potentially leading to conflicting obligations for businesses.

For example, the AI Act may mandate continuous automatic analysis of inputs and outputs, while GDPR could simultaneously restrict such automatic processing due to privacy concerns. These types of conflicts are already well-documented in the intersection of IT security and GDPR, highlighting the need for a harmonized approach that aligns requirements and mitigates conflicting regulatory demands. A cohesive, integrated strategy is necessary to ensure that compliance efforts are effective, streamlined, and practical for all stakeholders involved.

1.1.6 Question 18 Technical Risk Mitigation Measures Reflect Differences in Size and Capacity

See our response to Question 12 (1.1.2).

1.1.7 Question 19 Current State of the Art Specific Technical Risk Mitigation Measures

Measures should be implemented only to the extent that they provide clear added value, ensuring that regulatory requirements remain both effective and practical. It is important that such measures are foreseeable and well-defined; replacing ambiguous requirements under the AI Act with similarly unclear requirements under a Code of Practice (CoP) will not enhance regulatory clarity or compliance.

For example, content filters necessitate precise definitions of content types and clear expectations for filtering capabilities. Similarly, labelling mechanisms require commonly agreed-upon labels and standards for both attaching and analysing these labels to ensure consistency and usability across the board.

Given the short timeline, there is a strong case for prioritizing existing tools, elements, and internationally recognized good practices. Introducing entirely new approaches should be limited to situations where they are necessary, ensuring that efforts remain focused and manageable within the current regulatory landscape.

1.1.8 Question 21 Greatest Challenges for Implementation

See our response to Question 17 (1.1.5)

1.1.9 Question 22 Internal Risk Management and Governance Measures Reflect Differences in Size and Capacity

We suggest that regulatory frameworks should allow for functional, department-based approaches rather than placing obligations solely on individuals. This flexibility would enable organizations to leverage existing internal structures more effectively, promoting a more cohesive compliance environment.

Permitting the outsourcing of responsibilities to experts and providers with proven expertise is also crucial, as it allows organizations to tap into specialized knowledge and best practices, thereby enhancing the overall effectiveness of compliance measures. Additionally, enabling multiple obligations to be managed within departments or by individuals would facilitate alignment and reduce redundancies, particularly where overlaps exist with frameworks like GDPR, IT security, or business continuity.

There is significant value in considering existing good practices of internal governance, as seen in GDPR, IT security, and business continuity, as potential blueprints for structuring compliance under

the AI Act. By building on these well-established models, organizations can achieve more streamlined and integrated approaches to compliance that draw on proven methodologies.

1.1.10 Question 23 Internal Governance Measures

In reference to replies to Questions 19 (1.1.7) and 22 (1.1.9), it is crucial to avoid conflating different approaches when implementing measures. For instance, while a bug reporting program can be supported by bug bounties, the value of allowing bug reporting itself does not depend on the presence of paid rewards. In practice, it is important to address the issue that bug reporting is sometimes misinterpreted as hacking. This misunderstanding can place individuals who report vulnerabilities, despite their good intentions, at risk of criminal proceedings.

Furthermore, while certain measures, such as those related to IT security, may seem logical, it is important to tailor these measures specifically to the AI context. General IT security measures, like physical access control, should remain within the domain of IT security to avoid redundancy and potential conflicts with AI-specific requirements. Focusing on AI-relevant measures ensures that the regulatory framework remains coherent and avoids unnecessary overlaps or incompatibilities.

1.1.11 Question 25 Greatest Challenges in Implementing

Refer to our responses to Questions 17 and 21 (1.1.5 and 1.1.8)

1.2 Monitoring /Lifecycle Related Questions

1.2.1 Question 28 Review and Adaptation of the Content of the Code of Practice

Continuous evaluation of regulatory measures is highly valued, yet it is essential to maintain good practices in lifecycle management throughout this process. Clarifications and editorial updates can and should be made continuously to enhance guidance and extend adherence possibilities for CoPs.

However, more substantial, breaking changes should be implemented no more frequently than every three years. This timeline allows providers adequate time to adapt to significant changes. Any such updates should be introduced with sufficient advance notice to ensure a smooth transition.

KPIs and evaluations must remain realistic and reasonable, reflecting the identified risks without imposing undue burdens. While there is room for increasing rigor in evaluations, care should be taken to avoid measures that could hinder adoption or effective implementation. It is crucial to emphasize maintaining compatibility with evolving best practices in both European and international markets to ensure the CoP remains relevant and effective.

2 Free-Template Questions

2.1 Working Group 1: Transparency and copyright-related rules

2.1.1 General Remarks (Methodology)

See 2.5

2.2 Working Group 2: Risk identification and assessment measures for systemic risks

2.2.1 General Remarks (Methodology)

See 2.5

2.2.2 Art. 55 (c) keep track of, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, relevant information about serious incidents and possible corrective measures to address them;

To effectively manage serious incidents, a structured methodology should be established, following good practices from GDPR, IT security, and business continuity. This approach should include defining, qualifying, and managing serious incidents, with the possibility of extending elements to address AI-specific matters if they offer added value.

- **Step 1: Defining Serious Incidents**

Serious incidents should be characterized by their potential to cause significant harm or risk to societal matters. This contrasts with individual cases where less severe proceedings might be adequate. Key considerations for defining serious incidents include:

- **Significant Harm or Risk**: Incidents that have a considerable impact on society, as opposed to isolated individual cases or those where legal action may be limited.
- **Safety Concerns**: Situations with serious implications, such as life-threatening conditions or irreversible injuries.
- **Bias and Discrimination**: Incidents that affect societal groups broadly, rather than isolated instances which might be managed through individual legal proceedings.

- **Step 2: Criteria for Qualifying Serious Incidents**

To qualify serious incidents effectively, several criteria can be of interest:

- **Impact Assessment**: Evaluating the extent and nature of the harm or risk involved.
- **Severity Level**: Determining the seriousness of the incident.
- **Frequency and Recurrence**: Assessing how often similar incidents occur.
- **Affected Population**: Identifying the scope and scale of the affected group.
- **Regulatory Implications**: Analysing the potential regulatory consequences and requirements.

- **Step 3: Corrective Measures**

Implementing corrective measures should focus on both immediate and preventive actions:

- **Immediate Actions**: Define and enact measures that address current risks promptly and effectively.
- Preventive Measures: Develop strategies to prevent the recurrence of similar incidents in the future, integrating lessons learned into development cycles and general risk assessments.
- **Stakeholder Engagement**: Engage with stakeholders and share expertise, as is common in IT security, to enhance the effectiveness of corrective actions.

2.3 Working Group 3: Risk mitigation measures for systemic risks

2.3.1 General Remarks (Methodology)

See 2.5

2.4 Working Group 4: Internal risk management and governance for general-purpose AI model providers

2.4.1 General Remarks (Methodology)

See 2.5

2.4.2 Develop a methodology for internal risk management and governance:

To effectively maintain internal risk management and governance, it is crucial to establish a structured methodology that adheres to established good practices. This approach should draw from proven frameworks such as GDPR, IT security, and business continuity, while also considering extensions for AI-specific matters where applicable.

- **Risk Identification and Assessment:** A fundamental component of this methodology is a thorough risk identification and assessment process. Risks should be systematically classified into common and broadly adopted categories, including technical, operational, legal, ethical, and reputational. This classification helps in understanding and addressing the various dimensions of risk effectively.
- **Risk Mitigation:** Effective risk mitigation involves implementing a layered approach, including:
  - Preventive Measures, i.e. Strategies designed to avoid risks before they occur.
  - Detective Measures, i.e. systems for identifying risks as they emerge.
  - Corrective Measures, i.e. processes for addressing and reducing the impact of risks once detected.
- **Governance Structure:** A robust governance structure should be established, based on good practices in internal governance. This includes the separation of duties and powers to ensure clear oversight and accountability.
- **Internal Training/Awareness and Continuous Improvement:** To support and sustain effective risk management, it is essential to implement ongoing internal training and awareness programs. These initiatives keep staff informed about current risk management practices and emerging threats. Additionally, continuous improvement programmes should be developed to refine and adapt risk management strategies based on new insights and evolving challenges.

2.5 General Remarks and Methodology

SRIW shares the following considerations with the intention to promote a balanced, adaptable approach to CoP development and application in the context of general-purpose AI models. By ensuring these codes are practical, aligned with existing regulatory frameworks, and conducive to stakeholder engagement, the regulatory ecosystem can better support compliance and innovation without imposing undue burdens:

- **Unclear Effects of General Validity**: We are of the view that the concept of general validity remains ambiguous, as it does within the GDPR framework. In our perspective, if general validity is to be established, it should not automatically render Codes of Practice (CoP) applicable and obligatory for all providers.
- **Potential Shift of Legislative Powers**: Automatically making CoP obligatory could potentially shift legislative powers to private stakeholders, raising concerns about maintaining a balanced distribution of authority between public and private entities.
- **Competition Law Concerns**: We believe there are also potential competition law concerns with making CoP mandatory. Such an approach could pose challenges, particularly in the early stages of CoP development, potentially discouraging relevant stakeholders from contributing due to perceived or actual competitive disadvantages.
- **National Legal Requirements**: It is also important to consider the diverse legal landscapes of member states. Even if the AI Act implies certain consequences of general validity, additional legislative acts may be required at the national level to formally acknowledge the role of CoP in legal and regulatory decisions.
- **Guarantees for Compliance**: In this context, it would be valuable if CoP, when implemented, could provide assurances that conforming with them offers prima facie compliance with the relevant sections of the AI Act.
- **Good Practices in Methodology**: Regarding the development of CoP, we suggest following established good practices and methodologies. This could include adopting a targeted approach, avoiding overly broad or one-size-fits-all solutions, and aligning CoP with diverse regulatory frameworks to prevent conflicting requirements. Emphasizing the reduction of bureaucratic burdens, enhancing scalability, and promoting high adoption rates would be beneficial. CoP should also build upon existing good practices, ensuring that requirements are actionable and implementation-ready.
- **Extensibility and Modularity**: Extensibility and modularity in CoP design could be valuable, allowing core requirements to be complemented by potential extensions. Such an approach would foster innovation-readiness, ensuring a level playing field while allowing precise distinctions where needed.

- **Focused Objectives**: We suggest that CoP maintain a clear focus on specific challenges and objectives, allowing for various means of implementation as long as the intended outcomes are met. Providing non-binding guidance, examples, and good practices could further support stakeholders in implementing these measures effectively.
- **Prioritizing Practical Challenges**: Lastly, we recommend prioritizing practical, real-life challenges with the highest impact before addressing less significant or theoretical issues. By focusing on the most pressing challenges first, CoP can deliver meaningful results and drive effective compliance within the broader AI regulatory framework.

selbstregulierung
informationswirtschaft e.V.