[AIG 24: Case 1] Example Documentation of Implementation Guidance for the EU AI Act: a draft proposal to address challenges raised by business and civil society actors[1]

Nyasha Duri
Independent

Organised by
Charlotte Siegmann, Esben Kran

Abstract

The purpose of this paper is to provide an example of documentation based on a draft proposal for an approach to the implementation of the European Union (EU) Artificial Intelligence (AI) Act. Its overarching aim is to explore creative ways to effectively operationalise the legislation in its current form, informed by the present discourse gauging feasibility.

Within that, the specific objectives are integrating potential solutions to some of the key challenges that have been raised by business and civil society actors. Its scope is exploring generalisable recommendations for mechanisms that may help lead to successful implementation, geared towards public, private, and third sector leaders.

In summary, concerns primarily and predictably centre around challenges with AI governance needing to find elusive balance between minimising innovation impediments, and protecting people from what the EU terms 'unacceptable' outcomes at the furthest end of its scale - necessitating outright bans. Along this spectrum, another element of exacerbation is the high-risk nature of certain contexts with particularly impactful use cases.

[1] Research conducted at the AI Governance Research Sprint, 2024 (see https://alignmentjam.com/jam/governance)

I conclude that as a result, potentially underexplored mechanisms such as civic technology-enabled participatory governance should be considered for greater collaboration between public, private, and third sector actors. While there are limitations, going beyond consultation to implementation guidance co-design with weighted power-sharing in terms of decision-making can assist with solving some of the problems. I outline a proposal for a zero-trust system to be hosted by government bodies that will enable organisations with live products or planning to go to market to audit and adapt accordingly.

Keywords: AI Governance, safety infrastructure

1. Background

Research questions:

- What are some of the key challenges that have been flagged with implementation of the EU AI Act and by whom?
- Which (public) solutions already exist to assist organisations and are there any gaps that could be filled?

My work is intended to contribute by proposing a novel approach to implementation guidance. This would theoretically be operationalised in practice through partnership between the European Union and key actors in the space among those who will be most affected by the regulation. Specifically in terms of having to dedicate resources to becoming compliant or whose organisations focus on fostering accountability for the latter. This will be presented throughout in the form of a draft proposal with example documentation accompanied by explanations.

In summary, key takeaways from preliminary enquiry (AI Now, 2023; Gerlach, 2023; Kran, 2023, Siegmann, 2023, Wörsdörfer, 2023..) are that there are:

Select Challenges

- fears about the administrative burdens affecting startups especially
- misconceptions about who will be subject to it and at which stage
- concerns about blanket mis-applications to areas such as open source

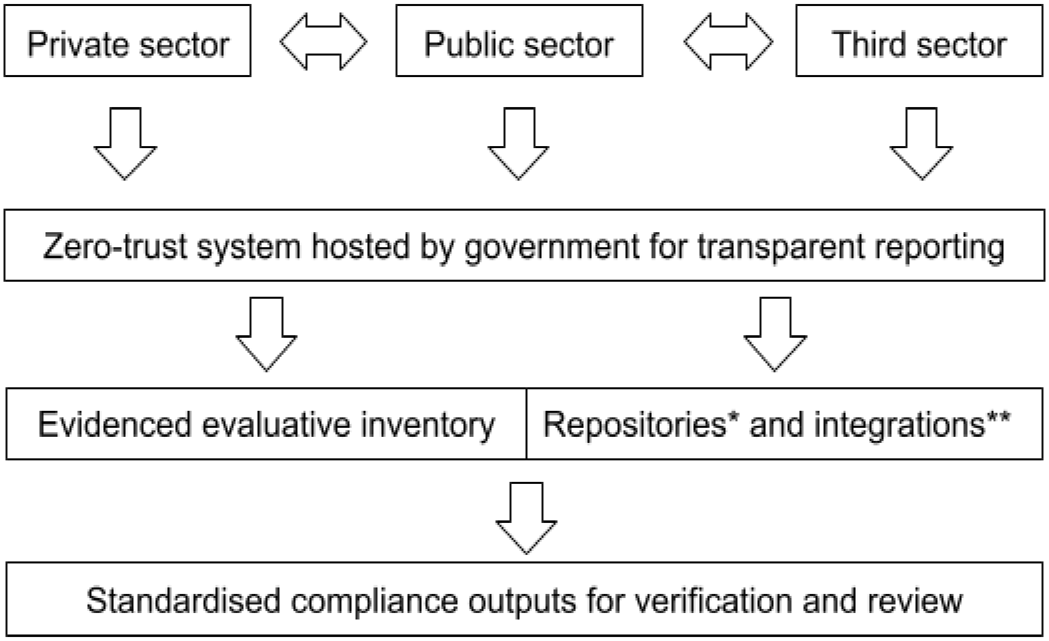
Potential Solutions

- a number of AI governance startups focusing on compliance support
- leading actors building their own self-regulatory preparedness frameworks
- gap in applying zero-trust models, freely accessible platforms and more

Lessons can also be learnt from how implementation guidance for the General Data Protection Regulation [EU 2016/679] (GDPR) cascaded top-down from the EU to member states, with the current landscape in terms of AI advancement necessitating approaches scaled up approaches to this extensive data ingestion processes involved.

An additional pendulum swings between accessible compliance for startups, and possibly entrenching arguably dangerous dominance of established actors in terms of data, compute, and other resources. That creates issues preferable to avoid such as single points of failure, highlighted by US, UK, Chinese, and other government bodies.

2. Draft Proposal


Private sector
Public sector
Third sector
Zero-trust system hosted by government for transparent reporting
Evidenced evaluative inventory
Repositories* and integrations**
Standardised compliance outputs for verification and review

*self-hosted
** real-time

The first aspect of the proposal is a participatory governance process with these 3 sectors co-designing a zero-trust system. The majority of outputs will be viewable by all, similar to existing business registry services.

While default decision-making will err toward democratic power sharing as much as is practicable, government will be the ultimate decision maker where there is disagreement.

In terms of mechanisms themselves, the current plan is for the co-designed system to offer multi-level interactive implementation guidance through a platform that records both mandatory and voluntary disclosures. Organisations with live products or ideally pre-deployment, for those newer in the space planning to go to market, will essentially undertake an audit that helps them adapt accordingly if relevant.

After answering questions to estimate their initial classification, building upon what the EU has outlined so far, they will be taken to an inventory list for them to follow along with at the base level. This enables uniform interpretation of the complexity of the legislation in a more accessible way. The goal is for this to be embedded into development processes, rather than (perceived as) an administrative burden.

At the secondary and tertiary levels, they will be required to provide evidence for their self-assessment, with permanently archived links to any additional resources self-hosted via their own or other means. As alluded to in the prior section, learning from states' approaches to implementation guidance for the GDPR applies here, with enhanced approaches needed to expand beyond e.g. checklist approaches that helped clarify compliance and flag breaches.

Here is what those implementation guidance fields and prompts could look like:

Inventory Item #-- and corresponding component	Detailed evidence for your assessment that can be verified	Additional links (e.g. to repositories: code, datasets etc)
□		

3. Limitations

Some limitations include difficulties with getting consensus, technical feasibility, as well as from an overall point of view the likelihood of a bias towards reactivity rather than proactivity. Particularly amid the importance of including third sector perspectives across the region, whose high likelihood of advocacy for the inverse of what incentivises the private sector may be a time-consuming blocker.

From a technical standpoint, it is unlikely to be viable to include real-time integrations e.g. API, always on communication channels, (if any) that would allow this to be more proactive - similar to managed detection and response protocols - rather than reactive. There would also be added cybersecurity risk vulnerabilities.

With more time, more specific examples would be explored.

4. Conclusion

In conclusion, there is a need for alternative approaches to AI governance that do not default to a combative approach between the private and third sectors especially - seeking opportunities for common ground. As the middle ground bringing those together in many instances, the public sector has an opportunity to empower the often under-served third sector more as the voice of ordinary people more often than not, while also better partnering with the private sector rather than being perceived as enforcing cumbersome burdens that impede innovation - while still putting safety first.

5. References

AI Now Institute (2023). Zero-Trust AI Governance. Sourced from: https://ainowinstitute.org/wp-content/uploads/2023/08/Zero-Trust-AI-Governance.pdf

Anthropic AI (2023). Anthropic's Responsible Scaling Policy, Version 1.0

Coalition for Health AI (2023). Blueprint for trustworthy AI implementation guidance and assurance

for healthcare. Sourced from: https://

www.coalitionforhealthai.org/papers/blueprint-for-trustworthy-ai_

V1.0.pdf.

European Union (2023). Texts adopted. Artificial Intelligence Act. Sourced from:

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

Gerlach, N.(2023). The case of the EU AI Act: why we need to return to a risk based approach. IAPP; Sourced from: https://iapp.org/news/a/the-case-of-the-eu-ai-act-why-we-need-to-return-to-a-risk-basedapproach/.

Kran, E. (2023). Spoken conversation

Siegmann, C. (2023). AI Governance Sprint Keynote

Wörsdörfer, M. (2023). AI ethics and ordoliberalism 2.0: towards a 'Digital Bill of Rights'. AI Ethics. Sourced fromL https://doi.org/10.1007/s43681-023-00367-5

(Appendix and other additions to follow separately as with presentation)