BRIEFING

STOA Options Brief

European Parliament

Regulatory divergences in the draft AI act

Differences in public and private sector obligations

This STOA Options Brief summarises the policy options developed in the accompanying study. The study identifies and examines sources of divergence in the obligations facing public and private sector actors when applying artificial intelligence (AI) in the draft AI act (AIA). It focuses in particular upon manipulative AI, social scoring and biometric identification

Table 1 – Diverging obligations for public and private sector users under the draft AI act

	Public sector	Private sector
Manipulative AI systems	Art. 5(1)(a),(b) AIA bans the use of manipulative AI systems in case of intentional harm and for a limited number of exploited vulnerabilities	Art. 5(2) of the Unfair Commercial Practices Directive (UCPD) protects against the commercial exploitation of 'average consumers'. Art. 5(3) UCPD protects vulnerable groups from unintentional unfair private sector practices
	Annex III, 6(c) AIA lists AI systems used by law enforcement to detect deep fakes as high-risk	Art. 52(3) AIA envisions only transparency obligations for deep fakes (low-risk)
Social scoring	Art. 5(1)(c) AIA bans the use of social scoring for public authorities	Social scoring for private sector actors (see Annex III AIA 3(a) and (b), 4(b), 5 (a), (b) and (c)) falls under the high-risk regime, Art. 6 ff. AIA
	Art. 5(1)(c) AIA – includes a prohibition of data sources for public sector ('social behaviour or known or predicted personal or personality characteristics')	Art. 10 AIA – (mere) data quality and governance requirements for private sector social scoring
Biometric AI systems (BIS)	Art. 5(1)(d) AIA bans the use of real-time BIS for law enforcement	Use of real-time BIS by private sector actors falls under the AIA high-risk regime, Art. 6 ff. AIA
	Art. 52(2)2 AIA does not require transparency obligations for law enforcement use of emotion recognition systems (ERS) and biometric categorisation systems (BCS)	Art. 52(2) AIA envisions transparency obligations for the use of ERS and BCS by private sector actors

1. Strengthen risk assessment and criteria

The draft AIA is different from other recent EU legislation, such as the General Data Protection Regulation (GDPR), in the sense that the normal risk-based approach to regulating a technology has been expanded from a procedural obligation to include more substance. The act as proposed includes a classification of risks in relation to AI by dividing applications of AI into three categories: prohibited, high-risk and low-risk. Although such a categorisation can aid in regulating certain applications of AI across the board, it remains to be seen how and through which arguments applications will shift between these three categories. The draft AIA's novel element – the combination of a risk-based

approach (which is intended to make the legislation technology-neutral) with a technology-specific risk classification regime – could lead to confusion and legal uncertainty. Some of the AI applications are put in a risk category due to risks of fundamental rights breaches (and are thus based on a fundamental rights principle), whereas others are placed in a category based on a self-performed risk assessment (and are categorised under a risk-based approach). There appears to be an indirect assumption that more complex AI systems present greater risks.

We have seen in many cases of actualised harm that this is not at all the case. Such assumptions about the risk of applying a particular technology imply that a regulation has to state something technology-specific: **it means we somehow have to compare, weigh and categorise types of AI applications beforehand**. Whereas the analogy with, for instance, dangerous chemicals can be drawn when it comes to introducing risk categories in a regulation or law, the major difference is that the harms of substances can be tested and measured; when it comes to AI, this is far more elusive. Moreover, the problem of these two approaches is that the more we move down the 'ladder' from principle-based regulation to self-regulation, the more responsibilities are put on private actors, who in most cases also develop the AI themselves. However, many AI-associated risks are indirect and likely to manifest themselves over time, even if the system is legally or formally compliant. It is therefore difficult to imagine how an ex-ante risk assessment obligation will adequately anticipate AI risks, and will effectively do so also for risks that go beyond financial or reputational ones (those on which private actors would primarily focus).

The draft AIA's risk categories are not always applied consistently when it comes to public or private actors, and their obligations to mitigate such risks vary. Following the argumentation that many AI systems blur the line between public and private spheres and will continue doing so – and in line with the GDPR as a risk-based regulation that does not make a clear distinction between these two – a policy option would be **to apply risk categories and the obligations they bring about more coherently within the draft AIA between public and private actors**. Moreover, both the risk-based approach and the risk criterion levels could be made more concrete. Examples in our analysis of biometrics and social scoring in the draft AIA show an inconsistency in applying criteria, begging the question of how the deployment by private actors is less threatening than public use of such AI applications, especially since the checks and balances present in democratic societies for public institutions offer more fail-safe mechanisms than forms of self-regulation delegated to the private sector.

When looking at external diverging effects related to risk assessment, **it could be made clearer what the risk assessment is precisely about, and to provide clear delineations or cut-off points**. The risks of algorithms and models cannot be separated from the risks of data, because the risk of an algorithm manifests only when it is acting on real-world data. Since the data used to either train or deploy an AI system is in itself already part of other risk-assessment regimes, the draft AIA could lead to contradictory risk assessments and confusion about which risk assessment takes precedence in a particular technology's development or deployment. For instance, there are overlaps and divergences between the draft AIA and Article 22 of the GDPR, which regulates automated processing and profiling (so also potentially forms of AI). **Providing guidelines on how the draft AIA risk assessment interacts with other risk assessment obligations** put forward in many of the EU regulations and directives that deal with digitalisation (e.g. the GDPR's Data Protection Impact Assessment) would be a step towards regulatory coherence.

2. Strengthen information and disclosure obligations

The draft AIA couples many of the low-risk applications, including deep fakes and BCS, with transparency obligations. However, **mere disclosure of the fact that an AI system is at work does not lead to better protection of rights**, and certainly not so against the spectrum of public and private AI providers and users. Lack of transparency has led to a legal response in cases revolving around the public sector's use of AI and has come from an organised action (not from a single individual). A case in point in that regard is the Dutch SyRi. The lack of transparency of the algorithm used in this state-

deployed risk-profiling system was reason for a judge to put a halt to the system (based heavily on the GDPR). The lack of transparency, however, had a clear link to the auditability of the algorithm's performance and was not the primary or sole argument in proving harm. The Court thus argued that given SyRi's 'high-risk' features and high level of intrusion, if the system is not performing it does not pass the proportionality test. However, if it cannot be known how the algorithm performs because it is not transparent, then the system's legal basis cannot be tested and therefore should be prohibited.

While the example above does show that a transparency obligation **can** lead to some sort of a legal remedy, it also clearly **demonstrates transparency's connection to rights enshrined in the GDPR and other sources of law**, as well as to **an established system of checks and balances for public actors**. Yet, such precedents and case law remain rare. Transparency in the form of an AI-registry, as some municipalities in Europe are doing, also means very little in terms of preventing harm or of providing forms of legal redress; it is meant as an aid in legal disputes and not as protection in itself. Notably, such registries in the private sector exist for research purposes alone – social media and other big digital service providers aim to protect their intellectual property. Even if individual harm resulting from an AI system were addressed by a citizen or consumer, the multiple steps between transparency and individual redress would include a) being aware that one is harmed (which in the case of manipulative AI systems, social scoring or covert biometrics is near to impossible); b) being able to both identify and separate the AI-part from other parts of a service or system that caused the harm; and c) being able to argue that even if consent was given or assumed, harm was caused that could not have been foreseen or anticipated, and to then quantify that harm somehow. All these **steps are to be taken by the harmed party, while the AI application 'users' as stated in the draft AIA, merely inform the citizen or consumer that such a system is involved**. Further, the obligation does not apply if the AI system is used to detect, investigate, prevent or prosecute criminal offences – an exception that is too broad to do justice to the presumption of innocence. Whereas the draft AIA was developed in the context of ensuring trustworthy AI development in Europe, the lack of redress mechanisms and inconsistent connection to fundamental rights in the draft AIA could actually decrease trustworthiness. Note here again the difference with the GDPR, which imposes numerous transparency and information obligations on data controllers in order for data subjects to make use of their data-related rights, including Article 21 (right to object) and Article 22 (right not to be subjected to automated decision-making or profiling).

Transparency in the draft AIA is not linked to a subjective right, and remains as such at the level of principle or policy aspiration. An option to overcome this would be to clarify and directly stipulate in its provisions **how GDPR rights and remedies are applicable to the addressees of AI systems**, especially when data rights are involved, and to further critically assess the connection between the draft AIA's transparency obligations and redress mechanisms by **strengthening information and disclosure obligations with withdrawal rights**. The latter would strengthen individuals' legal standing, and contribute to rebalancing the private sector's asymmetrical AI powers.

3. Consider co-regulation strategies and impact assessments

AI systems are both transformative and disruptive, and their impacts on governments, companies and citizens is growing increasingly complex. However, these impacts are neither equally distributed nor accounted for. **The draft AIA appears in some instances to be regulatory strong-arming the public sector and its AI systems, while neglecting risks and harms of equal intensity and distortion produced by private sector practices**. In regulatory terms, however, consolidation of certain practices confers great power. Once such power dynamics (and the behaviour underlying them) have been standardised, hierarchical modes of ordering and the ways they exercise control, solve conflict or ensure compliance are rendered ineffective, as the way in which rule-making and rule-taking work out in practice do not overlap.

The role of the private sector in exacerbating, by means of providing AI-based services or products, existing societal issues, as well as in structurally transforming the ways in which we understand and

relate to consent, privacy, accountability and distributional justice, to name just a few of the examples discussed in this report, **alters traditional dynamics of rights and interests**. Such transformations – conditioned by the consolidation of the AI-related market position of the few big tech companies already dominant in other areas of the digital economy, coupled with the companies' ever increasing infrastructural presence in essential (and increasingly AI-driven) services previously provided by the state – affect our social, economic and interpersonal lives in a radical way, and **call for an evaluation of the design and provisions of the draft AIA that places them in a more prominent perspective**.

In its current form, the AIA approaches governance by a combination of hierarchy (law-centric) and self-regulation (techno-centric). Obligations that are not clarified in a top-down manner are left to the industry standardisation bodies to articulate, which severs the channels of communication between industry and external stakeholders. More importantly, this approach focuses largely on the material features of AI systems, omitting to incorporate in a more consistent way the underlying socio-technical changes and impacts such systems might have on individuals and society. However, AI systems can be tackled from both perspectives – be both regulated and used as regulatory tools – **provided that the negotiated norms have been co-created by all stakeholders involved**, and focusing on the relationship between the properties of the AI system and the capabilities of the actor that determines how the system will be used.

The combination of a technology-centric approach with a law-centric one is something that needs monitoring if and how, that is to say in which sectors and through which obligations and oversight bodies, it will play out. The draft AIA does not provide clear measures to monitor such regulatory effects beyond the ex-ante risk assessment and perhaps 'by-design' approaches through regulatory sandboxes. **Measuring long-term effects and socio-technical changes as a result of using AI systems by means of ex-post impact assessments is currently lacking. The draft AIA can be re-evaluated in this fashion to consider** incorporating ex-post impact assessments to better grasp **the trajectory and distribution of AI systems**, including the factors that drive its proliferation in sectors. This would allow legislators, regulators and policy-makers to consider issues arising from AI systems, and the activities and roles of private parties behind those in a holistic manner.

This document is based on the STOA study 'Regulatory divergences of the AI Act: Differences in public and private actors' obligations when using AI systems'. The study was written by Ilina Georgieva, Tjerk Timan and Marissa Hoekstra of TNO, at the request of the Panel for the Future of Science and Technology (STOA), and managed by the Scientific Foresight Unit, within the Directorate General for Parliamentary Research Services (EPRS), European Parliament. STOA administrator responsible: Philip Boucher.

DISCLAIMER AND COPYRIGHT

stoa@ep.europa.eu (contact)

http://www.europarl.europa.eu/stoa/ (STOA website)

www.europarl.europa.eu/thinktank (internet)

http://epthinktank.eu (blog)