The AI Act: When Too Much Is Not Enough

Anna-Mari Wallenberg

Cognitive Science, Department of Digital Humanities, University of Helsinki

Abstract

The AI Act has been widely celebrated for its emphasis on fundamental rights. However, while the Act makes grand promises, it ultimately may achieve very little. The regulation spans over a hundred articles and nearly two hundred recitals. Despite its massiveness, the AI Act may not effectively promote the realization of human rights; rather, it merely mitigates associated risks.

1. Introduction

In the course of history, new technologies have often sparked anxiety[1]. The discourse around artificial intelligence has been particularly tense. Headlines describe AI as a disruptive technology, which discriminates, exploits and manipulates. It destroys democracies, spreads misinformation and threatens humanity.

These concerns have served as catalysts to regulate AI. In April 2021, the European Commission ("EC") unveiled its drafts for the Artificial Intelligence Act ("the AI Act"). After years of preparation, the AI Act was approved by the EU Member States ("Council") and the European Parliament ("EP") in the spring of 2024. The Act came into effect in August 2024, with a transition period that ends in 2027.

The AI Act governs the development, commercialization, and use of AI products within the EU territory. It mandates pre-market requirements, including fundamental rights impact assessments and transparency obligations, on both users ("deployers") and developers ("providers", "manufacturers", "importers") of AI systems. Non-compliance will be met with significant fines.

[1] MacNaghten & al., (2015), Mokyr & al., (2015).

One of the AI Act´s main objectives is to address the fundamental rights concerns arising from the opaque use of AI (COM2021). However, it struggles to achieve this goal. As Almada and Petit (2023) remarks, there is a fundamental "mismatch between means and ends in the AI Act." Technically, the AI Act tries to extend the product safety regulation to cover fundamental right issues. This creates a set of challenges, which the AI Act does not fully tackle.

In this article, I will focus on these challenges. First, I will address the difficulty of crafting a future-proof, technology-neutral definition for AI (Sections 2-3). Next, I will analyze the AI Act's struggle to provide an account of fundamental right risks (Section 3), and the effectiveness of its risk mitigation strategies (Section 4-6). I´ll conclude by arguing that the AI Act falls short of achieving its objectives (Section 6). Namely, the AI Act promises much but ends up doing very little.

2. Challenge (1): The Scope of the Regulation

The fundamental challenge for any AI regulation is to determine the appropriate scope of legislation in a future-proof way. The rapid and unpredictable evolution of algorithmic technologies makes regulations tied to specific programming techniques quickly obsolete, leading to discriminatory legal practices (Greenberg, 2016)[2].

To avoid this problem, the European Union (EU) decided to regulate the risky *uses* of AI applications, not the technology itself. The solution would both ensure future-proof regulation, and be technology-neutral, preventing discriminatory practices.

In a more detail, The AI Act was designed to impose pre-market obligations ("ex-ante approach") on key stakeholders in value chains. The obligations are contingent upon the roles of stakeholders in designating the (intended) uses of AI systems, not on the used technology per se. In practice, the AI Act imposes responsibilities to the "providers" (e.g developers, such as software companies), and the "deployers" (e.g. users, such as a public sector agency using AI system for service-production)[3].

For example, if *a public sector deployer*, such as a hospital, employs an AI system for assessing the patient´s need for medical care, then the hospital decides what the concrete context, manner and domain of the use are. Under the AI Act, as a

[2] In juridical context, technology neutrality means that laws promote future proofness (´statutory longevity`), and equal treatment of old and new technologies. It prevents legal discrimination, as it enables the stakeholders to decide itself what technology is better suited for achieving their goals (Greenberg 2016; Shadikhodjaev 2021).

[3] There are also specific responsibilities to other actors, or stakeholders in the supply and value chains (e.g manufacturers, importers, and so on).

public sector deployer the hospital is required to conduct the mandatory Fundamental Rights Impact Assessment ("FRIA") before implementing the system in practice (art. 27). The FRIA obligates the deployers to examine how the *use of AI system* effects fundamental rights, ensure the human oversight, and to develop methods for effective risk-mitigation.

Since the FRIA assessment necessitates information about the technical details of the used AI system, the *provider of the AI system* (e.g. a software company that develops and brings the AI system to the market), must also furnish adequate information to the deployer (the hospital using the system)[4]. The provider designates what the intended purpose of the developed system is (Art. 3.12).

The required information must adhere to standards, necessitating its provision in the form of *pre-market conformity assessments* (Art. 8-15). Mandatory assessments, thus, pushes *the providers* to share information about the used training methods, the used data, as well to spell out the way how the AI system is supposed to work, what its intended purpose of use is, and what the potential risks are (Art. 8-15, AIACT2024).

Importantly, the logic of obligations ensures technology neutrality, as it targets on stakeholders and their roles in the use of systems. However, the ad hoc provisions of the so-called "general purpose AI" ("GPAI") articles (Art. 52a-e) turned the original approach on its head.

Initially, the GPAI articles were introduced during French Presidency (2022). The goal was to ensure the fair regulation of models capable of serving multiple purposes. The articles, however, were revised several times in response to the political pressure caused by the rise of generative AI in 2023[5].

After painfully confusing and difficult negotiations[6], the EU ended up imposing obligations for the providers of *general-purpose AI models* (such as GPT-4) and

[4] According to Art. 3.12, the provider defines the intended purpose of a use for an AI system.

[5] In particularly, they were reinterpreted in terms of foundation models. For the notion of "foundation model", see Bommasani & al., 2021.

[6] In the AI Act, the notions of "foundation model" and "frontier models", the rise of generative AI, and the regulation revolving around the intended purposes are lumped together. As a result, the distinction between GPAI-models and GPAI-systems is conceptually and technically confusing. For example, the general-purpose models are defined in non-technical terms of "tasks", and the GPAI-systems are defined in terms of "purposes", but tasks or purposes are not given any specification. This non-technical, conceptually confusing classification makes it hard to delineate, what the GPAI-models and systems are supposed to be. The Commission is preparing the so-called "codes of conduct" for helping stakeholders to interpret the regulation in practical matters. Unfortunately, they focus only on practical requirements, not fundamental questions related to the scope of regulation. For the drafts, see: digital-strategy.ec.europa.eu/en/library/third-draft-general-purpose-ai-code-practice-published-written-independent-experts#email

general purpose-AI systems (such as chatGPT). The GPAI- models are models, that “display significant generality and is capable of competently performing a wide range of distinct tasks” (art. 3.63), while the GPAI- systems are “based on a general-purpose AI model” and have “the capability to serve a variety of purposes (art. 3.66, art. 51)[7].

Providers of GPAI-systems are expected to provide technical documentation about the models, describe the content used for model training, and explain how they implement existing EU copyright law. The providers of advanced large-scale GPAI models with systemic risk must assess and mitigate systemic risks by performing model evaluations, keeping track of, documenting, and reporting serious incidents, and ensuring cybersecurity protection for the model and its physical infrastructure (Art. 55).

Although GPAI-articles are not explicitly technology-dependent[8], they are likely to steer the enforcement of the regulation toward such a direction[9]. For example, the recital 98 states how “the generality of” a GPAI model could… “be determined by a number of parameters” and that “models with at least a billion of parameters and trained with a large amount of data using self-supervision at scale should be considered to display significant generality”, as they “competently perform a wide range of distinctive tasks”. Even if this does not explicitly refer to any specific models, it clearly switches the focus from the uses of models to the technology itself.

GPAI- articles, thus, extend the scope of the AI Act to areas beyond the reach of the original regulation. The articles alter the logic of the AI Act, raising concerns on the AI Act´s internal coherence. Inconsistencies also add layers of ambiguity, which may impact negatively on the AI Act´s practical compliance.

3. Challenge (2): The Appropriate Definition of AI

Conceptual discrepancies impact on the legal certainty, endangering the effective enforcement of regulation. To avoid this, the regulation should articulate its key concepts with sufficient clarity. Key concepts are crucial: They delineate what falls within the scope of the regulation, and perhaps more importantly, what does not.

[7] Concretely, providers are expected to provide technical documentation about the models, describe the content used for model training, and explain how they implement existing EU copyright law.

[8] One may argue that GPAI-articles do not explicitly discriminate certain programming techniques, or model architectures, as they focus on models and systems with multiple purposes and tasks.

[9] The explicit formulation of the articles may not be technology dependent. The classification criteria (art. 51) raises a possibility of technology-dependent practices.

In the AI Act, the AI is one of the central concepts. After multiple proposals, the EU adopted the so-called OECD definition of AI. Strategically, the co-adoption of terminology with an influential economic forum is understandable. It may increase the global visibility and impact of the AI Act, facilitate the global harmonization of AI governance, and diminish the potential negative effects of the AI regulation on the EU´s competitiveness.

The OECD definition[10] characterizes AI as "systems" rather than focusing on programming techniques or software taxonomies. This makes the definition better suited for technology-neutral regulation. It adds flexibility, as it allows the AI systems to encompass multiple technical solutions (Art 3.1.):

"'AI system' means a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments"

However, the notoriously complicated definition prioritizes future proofing at the expense of clarity. The terms such as "may," "varying levels," and "can" are intended to tackle the rapid development of technology. They make the definition complex and ambiguous. Technical jargon ("predictions", "recommendations", "generate"), unnecessary clarifications (e.g. "physical or virtual"), and undefined everyday concepts (e.g., "content," "influence") blur the content more.

The core of the definition, however, is that AI systems are *autonomous* and *adaptive*, and they can *infer* (recital 6 and 12, and guidelines 2.20)[11]. Such adaptive systems have "self-learning capabilities" that allow them to "change" while they are in use (recital 12). Moreover, they are "*autonomous*" (guidelines 2.20), if they have "some degree of independence of actions from human involvement and of capabilities to operate without human intervention" (ibid.).

Thus, the definition *excludes* automated, non-autonomous, rule-based mechanical "traditional software systems" from the scope of the regulation (guidelines, section 2). As the recital 6 states, the concept of AI systems "should not encompass systems that are based solely on rules defined by natural persons to automatically execute operations". If systems are neither adaptive (as they only mechanically execute the pre-defined rules without learning) nor autonomous (as automation does not imply autonomy[12]), they are not AI systems (in a light of the article 3.1.).

[10] It defines AI as "*a machine-based system that can, for a… set of.. objectives, make predictions, recommendations, or decisions influencing real or virtual environments*" (OECD 2024).

[11] The Commission published a draft of guidelines for facilitating the interpretation of the definition. For the guidelines, see: digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application.

[12] For discussion, see for example Nersessian et al., 2021.

Is the Definition Too Narrow?

For critics, the exclusion makes the scope of the AI Act too narrow. For instance, Haataja and Bryson (2023) argue that "all software" that make "life-changing decisions" should be subject to AI regulation[13]. For Haataja and Bryson (ibid.), it "does not matter", whether the decisions are made by "excel macros, steam-punk clockworks" or complex machine learning algorithms.

It may well be that any software used for decision-making have fundamental right effects. The AI Act, however, is hardly the most suitable instrument for addressing them. The AI Act is (or, was intended to be) a fine-tuned instrument for preventing a (presumably) specific set of scalable risks raised by certain algorithmic systems (guidelines 2.20-21).

The use of adaptive, computationally complex, indeterministic and unpredictable algorithmic systems for "autonomous" decision making raises these risks in a way, that is anticipated to escape the human oversight (COM2021, guidelines 2.21). In the logic of the AI Act, the risks associated with the use of AI, are (or should be) conceived as systematic, typically scalable *side effects* of the computational profile of opaque systems (i.e. for example systematic vulnerability of well-trained and well-functioning models due to their computational profile[14]), their training history (e.g. data bias), when these systems are used for specific purposes.

Importantly, these systematic "side effects" are due to the computational profile of sufficiently complex systems (for example, brittleness is associated as a side effect of optimization of accuracy, see Ilyas et al, 2019), and their training and programming methods (data bias in the context of typical ML-based classification algorithms).

More simple programming techniques do not constitute, cause or raise similar problems, as their computational profile is different. This is not to say that excel macros or clockworks don´t have any negative, scalable and opaque human right effects. They may, but typically by "simple" performance errors, bugs, or bad design choices, and not by computational profile or its systematic fragilities[15].

[13] Haataja and Bryson (2023) analyses the definition of AI in a light of the notion of "intelligence", arguing that we should adopt the widest possible definition for intelligence that would cover all software. As I see it, there is no need to refer to the notion of "intelligence". The notion of computational profile suffices for the purposes of Article 3(1) and Recital 6.

[14] For example, the systematic sensitivity to adversarial features, see Ilyas et al., 2019.

[15] For an analysis of constitutive, causal and algorithmic analysis of computational systems, see Craver 2000, Rusanen & Lappi 2007, 2016, and in the context of contemporary LLMs Millier and Bucker, 2024.

To sum up, it does not make sense to extend the required, relatively massive, tailored pre-market risk assessment methodologies (Art. 8-15)[16], such as mitigation of adversarial attacks (Art. 15(5)), to cover *any* software. It suffices, if it regulates the relevant ones[17].

4. Challenge (3): The Legal Legacy of Prior Product Safety EU Regulation

The legacy of EU product safety regulation plays a significant role in the challenges faced by the AI Act. The AI Act aims to reform EU product safety regulation by extending it to cover fundamental rights. However, this leads to problems that the AI Act does not fully resolve.

In line with the established approach of existing EU legislation, the AI Act categorizes the uses of AI based on the nature and severity of the risks they pose. Following the canon, the AI Act defines risks as “the combination of the probability of the occurrence of harm and the severity of that harm” (Article 3.2). The “probability of occurrence” refers to the likelihood of a harmful event[18]. “Harms” are violations of fundamental rights (such as safety, health, or equality), threats to “European values,” or acts against the “public interest” (recital 5).

Harms may be “material or immaterial, including physical, psychological, societal, or economic harm” (recital 5). Certain harms are systemic, manifesting as large-scale adverse impacts on public and economic safety, security and health, critical infrastructure, or democratic processes (recital 110). Others may affect individuals (recital 50) or groups of persons (recital 30).

The classification of risks forms a taxonomy of four risk levels, namely uses with 1) unacceptable risks, 2) high risks, 3) limited risks, and 4) low and minimal risks.

Prohibited AI practices. The intended purpose of a system use, which raise *unacceptable risks* contradicts EU´s “fundamental values”, such as respect for human dignity, freedom, equality, democracy, fundamental rights, and the rule of law will be prohibited (art. 5, recital 28). For example, social scoring, untargeted scraping as a form of privacy and data right violation, algorithmic exploitation of vulnerable

[16] There is no systematic, independent and reliable impact assessments on the short run or long term costs, or effects, of the conformity assessment, or the FRIA-methodology at the level of individual EU member states.

[17] One should be ready to evaluate the ideal of technology-neutrality objectively. As Greenberg 2006 emphasizes, in some occasions technology-specificity may serve better the regulatory purposes.

[18] The AI Act does not define the “probability of occurrence”. The interpretation is based on Novelli & al. (2024) interesting paper on the static notion of risks.

groups, and the unauthorized surveillance use of real-time biometric identification in public places for the purposes of law enforcement are prohibited (art. 5).

High-Risk uses. If the intended purpose of a system uses poses "high-risk" (Art. 6), the use is authorized, but subject to requirements and certifications to access the EU market (following a premarket conformity regime, "conformity assessment")[19]. The high-risk uses cover the 19 cases listed in Annex II, for which there is already existing legislation (such as medical devices, machines or toys), and the additional eight use areas listed in Annex III (such as migration, education, public services, critical digital infrastructure, and medical aid).

Limited risk uses are, for example, situations where AI systems interact with humans, such as chatbots or deepfakes in consumer services. As the developed deepfake techniques make it possible that humans may not understand they are communicating with machines, to prevent this risk the AI Act subjects light transparency obligations, such as codes of conduct, to these uses (Art. 50).

"**Low or minimal**" **risk** uses pose negligible or no risks to individuals' rights, safety, or well-being, and therefore, no direct regulatory requirements are imposed on them.

Curiously, the AI Act does not specify, *why* the risk-levels are as they are, or what, exactly, are the criteria used to estimate the type, quality, or severity of assumed harms[20]. Instead, the AI Act provides lists of fundamental rights (recital 48) and it *stipulates* that certain uses of AI raise risks to these rights[21].

This "stipulation-based" (my term) approach characterizes the risk profile of a high-risk AI system as a combination of the two dimensions[22]:

(i) if a system is an AI system (the definition of AI in Article 3.1), and

[19] The obligations are designated primarily to the providers (=suppliers, such as a software company) and the deployers (e.g. a public sector agency using AI in service production) of AI systems.

[20] The FRIA assessments and conformity assessments push the stakeholders to analyse, how the use of AI systems utilize personal data, what are the security and cybersecurity issues raised, and whether the use raises questions, say, on the equal treatment of people, human dignity, or other rights. They operationalize fundamental right risks as the potential effects that the uses of opaque, scalable, bias-sensitive and technically brittle AI systems may cause to human rights, such as equality rights (such as the discriminatory uses of biased AI when assessing creditworthiness), human dignity and autonomy (such as the algorithmic manipulation of sensitive groups), and privacy, data protection and cybersecurity (Art. 8-15, 27, recital 56).

[21] Lists in Annex II and Annex III covers the high-risk cases, and Article 5 the prohibited uses.

[22] Correspondingly, (i) if a system is an AI system (say, a machine vision system), and (ii) the use area (say, emotional recognition from facial expressions) is mentioned in the Article 5, and there are no exceptions that would allow the use of systems (specified in the article 5), the use of an AI system in that use case area is categorically prohibited.

(ii) if the use area of the system is mentioned either in the lists of eight high-risk use areas (Annex III) or in the list of 19 cases, for which there is already EU legislation (Annex II),

(iii) then the use of a system is a high-risk use, which raises risks to fundamental rights.

Originally, the Commission motivated this approach by stating that it establishes "predictable, proportionate and clear obligations" for the providers and developers of AI systems (COM2021). Indeed, it proposed a straightforward procedure ("look at the lists in Articles and Annexes") for determining whether the use of AI systems falls into the category of high-risk uses, or not.

The proposal, however, is problematic. It does not articulate, why certain uses are classified as high-risk (Annex III) or prohibited (Art.5), while others are not. For instance, one might ask why the algorithmic safety components of critical infrastructure (AI Act, Annex III, point 2) are deemed to pose risks to fundamental rights, whereas AI systems used for pricing energy consumption are not.

Moreover, the AI Act does not explicate the metrics used to assess the levels of risks. This invites to ask, why, exactly, for example AI systems intended for evaluating the creditworthiness of natural persons (Annex III, 5.b) are classified at the same risk level as systems designed to influence the outcome of an election (Annex III, 8.b).

The lists of selected risks are fragmented and piecemeal. Some of high-risk cases are chosen, as there is already product safety legislation covering them (Annex II). For some, the reason may be the presence of equality issues (e.g. AI-based discrimination). Some revolve around the privacy and data protection (e.g. the effect of GDPR), and some are reactions to potential economic, physical or psychological harm.

Generally, the stipulation-based approach illustrates how difficult it is to integrate the fundamental right issues with the product safety regulation. Philosophically, the rights do not follow the logic of product safety regulation. For a product safety regulation, it suffices that the systems keep the "risk-level below the bar" (Almada and Petit, 2023). In contrast, fundamental rights are a matter of principle, and they follow a "logic of optimization" (Almada and Petit, 2023). Fundamental right thinking aims to promote rights, not just mitigate its violations.

5. Challenge (4): The Legal Legacy of Prior EU Regulation (GDPR)

The legacy of prior EU regulation cause also other challenges for the AI Act. For example, the article 6.3 is designed to prevent the overregulation of AI systems in high-risk use areas. The article excludes assistive and procedural uses, emphasizing that only when AI systems[23] play "substantial role" in "decision-making" and "materially" influence "the outcome of decision-making" (rec. 53), they should be classified as high-risk uses. For example, the use of AI for narrow administrative tasks, classification of documents, structuration of the unstructured data, and the use of AI to "refine language" are not classified as high-risk (rec. 53)[24].

From a practical point of view, the exclusion of assistive uses is logical, important and valuable. It complements the definition of AI (art. 3.1), focusing only on the uses of AI, where machines are used to make decisions autonomously.

Unfortunately, the conceptual structure behind the current formulation of the article 6.3 is somehow problematic. Namely, following the GDPR (art. 22), the article 6.3. categorically classifies the complex computer-assisted and distributed decision-making as human decision-making (recital 53). It supposes that machines raise significant fundamental right risks *only* when they make the decisions, are used for profiling (Art. 6.3(d)), or have "substantial" role in decision-making (recital 53).

According to growing empirical evidence, this distinction between a man and a machine is an outdated idealization. From a cognitive science point of view, it ignores a very significant source of risks: the human cognitive factors, and their impact in distributed forms of machine-human interaction[25].

For example, the assistive uses of cognitively active AI technologies is found to carry increased risk (Slattery et al., 2024). These technologies, such as algorithm-enhanced measurement and analysis devices, interactive predictive data-analytics, or interface-integrated multi-modal assistants, are already widely used to support and extend human epistemic and cognitive capabilities in knowledge intense fields such as healthcare[26], science[27], governance and policymaking.

Such technologies may exert persuasive influence, directing users cognitively toward certain interpretations and actions (Weinmann & al, 2016; Mele & al, 2021; Wenker 2022). Moreover, studies on automation and other cognitive biases suggest that people are inclined to have excessive reliance to proposals generated by AI, even in situations where they have a reason to believe that information provided by

[23] There are some exeptions. For instance, if AI systems are used for profiling, then they are high-risk (Art. 6.3(d)).

[24] The AI Act demands that a provider who considers that an AI system is not high-risk should reliably document that the use of a system is narrow, assistive and procedural to avoid loopholes. Moreover, to ensure "transparency" and "traceability" that documentation should be provided to national competent authorities upon request (Recital 53).

[25] I think the article 6.3 is very valuable, and it should have been modified to exclude explicitly a narrow set of assistive uses in epistemic tasks (with direct fundamental right implications) rather than "material" uses.

[26] for example, Mele & Russo-Spena, 2019.

[27] For an overview on the impacts of AI in science, see Gao & Wang 2023.

the systems is likely to be incorrect[28]. In particular, certain user groups, such as semi-advanced novices, are found to exhibit such over-reliance (Horowitz & Kahn, 2023).

The increasing reliance on such technologies pushes us to recognize, how the cognitive dynamics of interaction challenges the fundamental assumptions of prior regulation. In many expert tasks, the human decision-making *is* assisted, technology-mediated decision-making, where the risks are raised by the interaction, not by the machines per se.

By relying on idealizations that do not recognize the changed cognitive dynamics, the regulation is deemed to view the sources of risks trough distorted lens. Reforming such fundamental commitments of the EU regulation would be a massive challenge to law makers. Still, in a light of the growing epistemic impact of cognitively active technologies, there may be a growing pressure for such a reform.

5. Challenge (4): The Effectivity of Transparency

The prior EU regulation, and particularly GDPR, have heavily influenced also on the AI Act´s risk-mitigation strategy. In short, following the legacy of GDPR, EU regulation tend to propose transparency and human oversight as "blanket remedy" for almost any kind of problem that the use of AI may raise (Ruschemeier & Hondrich, 2024).

However, neither transparency nor human oversight automatically ensures the effective prevention of fundamental rights risks (Ott et al., 2022) [29]. Decision-making can be fully transparent and controlled, yet rights can still be violated (Rodrigues, 2020).

Moreover, in the AI Act transparency serves mainly *communicative purposes* (Busuioc et al, 2023). It mainly facilitates the transfer of information (Busuioc et al, 2023; Mylly 2023), offering a means to overcome information asymmetries between stakeholders in the value chains (Mylly 2023).

In a more detail, under the Articles 8-15, any high-risk AI system must be transparent enough to allow deployers – the ones putting an AI system into use under their own authority - to interpret its outputs and use systems 'appropriately'[30]. The mandatory *conformity assessments* (Art. 8-15) aim to ensure that the providers pro-

[28] For discussion on automation bias in legal contexts, see Ruschemeier & Hondrich 2024.

[29] For discussion, see Daneshjou et al., 2021, Vaassen 2022, Viljanen 2023.

[30] For example, a provider must offer sufficient information (data management, technical reports, possible risks associated with the intended and unintended uses of systems, and risk mitigation plans) to deployers.

vide sufficient information about the ways, how the systems are trained, what algorithms they use, what risks they carry, and how they work[31]. Additionally, providers of systems are obligated to register high-risk AI systems in a publicly accessible, EU-wide database before these systems are marketed or deployed.

Transparency plays a similar, limited communicative role in *The Fundamental Rights Impact Assessment* ("FRIA"). The FRIA requires public sector deployers, such as government agencies, to detail the intended use of the high-risk AI system, including its duration and frequency (art. 27). The FRIA mandates identifying the categories of individuals or groups likely to be affected in the specific context and estimating the specific risks of harm these groups might face. The results of FRIA-assessments will be reported to national authorities, and light summaries of assessments will be registered in public EU database[32].

Since the FRIA assessments rely on technical information about the AI system, providers (e.g., software companies) must supply adequate information to deployers. The AI Act compels providers (such as companies producing software for hospitals) and manufacturers (e.g., software companies or research laboratories) to be more transparent by requiring that they share details about training methods, data used, and the system's intended functionality to the deployers through the conformity assessments (art. 13).

In a sum, the AI Act requires, thus, that the AI implementation complies with the relevant standards[33], that sufficient oversight methods are specified, and that the obligations concerning the registerations in an "open" EU-wide database are executed appropriately[34].

[31] In a more detail, the AI Act requires providers, such as software developers, to describe technical and data management details, identify potential technical vulnerabilities, estimate risks associated with both intended and unintended uses of the systems, and develop risk mitigation plans for high-risk systems (Art. 9). These plans aim to ensure compliance with fundamental rights and specify measures to address materialized risks. Providers must also explain the design and implementation of human oversight measures.

[32]https://www.euractiv.com/section/artificial-intelligence/opinion/eus-much-heralded-ai-act-agreed-by-eu-parliament-but-serious-human-rights-holes-in-law-remain/

[33] What comes to the AI standards, their legitimacy is also compromised (Mylly 2023). The standardization processes should be open to participation. In practice, it has been executed by a limited number of organizations and entities. There is no evidence on active promotion of participation, or that relevant civic society organizations would have been consulted (Mylly 2023).

[34] Under the article 86, "affected persons" also have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken. This extends the notion of communication transparency to "right to explanation".

The efficiency of the apparatus for increasing transparency is, however, an open question. First, the AI Act enables the providers *themselves* to assess, whether the technology meets the transparency and quality requirements. This may create "an increased risk" for "perpetuating sub-optimal transparency practices at the level of evaluation and control mechanisms" (Busuioic et al., 2023).

Second, the public database will contain basic information about the uses of systems, information on the providers of systems, and the EU declarations of conformity in cases where there has been self-assessment. (In the case of third-party assessment, the database will contain information about the certificate issued by the notified body.) The database will thus mostly reveal, which systems are registered. Once carried out and reported on, the database itself will not have any further, direct impact on the actual use of high-risk AI systems.

What comes to the mitigation of the fundamental right risks by pushing the deployers to evaluate the risks by the FRIA procedure, it is worth for noting that the FRIA is limited to the public sector bodies and private bodies carrying out public sector tasks in high-risk use areas (Art. 27, and Art. 50)[35]. Commercial "deployers", such as companies using AI systems for service production, are not required to perform FRIA assessments[36].

In a light of the increasing effect of commercially produced and distributed algorithms, this asymmetry between public and private sector obligations is unexpected, and perhaps slightly disturbing[37]. Commercial AI-products are associated with a growing number of harms and hazards (for example, the OECD AI Incidents Monitor 2024)[38]. Moreover, in many EU countries, public administration is already subject to very strict regulation, and human rights are safeguarded by constitutional legislation. Thus, one may ask, why the EU uses product safety regulation to regulate constitutional questions.

[35] The FRIA methodology continues the techniques of PIA ("Privacy Impact Asssesment") and DPIA (data protection impact assessment). They were introduced with the General Data Protection Regulation (Art. 35 of the GDPR). PIA/DPIA are associated with the obligation of the controller to conduct an impact assessment and to document it before starting the intended data processing.

[36] All deployers of high-risk AI systems that are used the detection of financial fraud (Annex III, 5b) and risk assessment and pricing in the case of life and health insurances (Annex III, 5c) must also complete the FRIA assessment.

[37] Market surveillance authorities, by virtue of article 47 of the Final Provisional Text, may lift the obligation to conduct a FRIA in case of "exceptional reasons of public security or the protection of life and health of persons, environmental protection and the protection of key industrial and infrastructural assets". Even in such case, article 47.1 of the Final Provisional Text states that this derogation is only temporary and completion of the procedures (e.g. conformity assessment and the FRIA) shall be undertaken "without undue delay".

[38] https://oecd.ai/en/incidents

The explanation is, again, the background of the AI Act. Human rights treaties tend to focus only on the government-citizen relationship, and their relation to corporations or companies is indirect (Engenström 2022). Traditionally, the legimated division of power between the state and a citizen raises the need for the special protection of citizen rights, and thus, states are conceived as the prime bearers of rights and duties.

As a result, the AI Act does not fully recognize the role of commercial actors for fundamental rights. In a light of the growing and undisputable economical and practical dominance of tech giants this, of course, illustrates the limitations of the available regulatory techniques. For an EU product safety law, however, it may not be a realistic option to ask for the regulatory reform at this level. Again, this indicates how complicated it is to balance between the available regulatory techniques with the realities of a tech-dominated era.

6. Concluding remarks: Too much is not enough

The AI Act represents a significant step in the regulation of artificial intelligence. It, however, struggles to achieve its main goals. Many of the AI Act´s challenges are due to a “fundamental mismatch between means and ends” (Almada and Petit, 2023). For a product safety law, it suffices that it keeps the “risk-level below the bar”, the systems meet the minimum standards and satisfy the requirements (ibid.).

Philosophically, fundamental right regulation would require more. It follows a “logic of optimization” (Almada and Petit, 2023), protecting and promoting rights to the maximum. This would require additional regulatory actions, such as significant and tangible legal incentives, equal access to participation, and regulatory inclusion of other democratic processes[39]. Moreover, the legal norms should promote the interests to which the right in question is directed in practice (Campbell, 2006; Griffin, 2004).

In the final form, the AI Act comprises over a hundred articles, and nearly two hundred recitals. The length of complementing guidelines is already over 150 pages, and more is to come. Despite its massiveness, the AI Act regulates only risks. It conceives human rights only in terms of harms, risks and risk-mitigation, without seeking any possibilities.

The legal legacy of the AI Act limits the available regulatory techniques and poses boundary conditions. Moreover, all the discrepancies, inconsistencies and provisions weaken the AI Act´s impact, and its cost-efficiency is an open question[40]. The AI Act may, thus, make great promises, but ends up doing very little. Simply put, sometimes too much isn´t still enough.

[39] For discussion on the protection vs. promotion of basic rights, see for example Shue 1996, Nussbaum 2006, Wibye 2022.

[40] As things stand, there are no reliable estimations about the costs of the compliance, or about the cost-efficiency of the whole exercise.

Acknowledgements:

I am grateful for my colleagues in my previous posts in Finnish Governance for many inspiring and thought-provoking discussions on the details of the AI Act: Olli-Pekka Rissanen, Riitta Autere, Maria Hauptmann, Kristine Alanko, and the members of the Finnish National Expert Group on AI and digitalization: Heikki Mannila, Petri Myllymäki, Sasu Tarkoma, Laura Ruotsalainen, Jukka Riekki, Minna Ruckenstein and many more. In particularly, I would like to thank the amazing Antti Helin for the years we worked together on the AI Act. Without those years this paper would have been so much less.

References:

Busuioc, M., Curtin, D., & Almada, M. (2023). Reclaiming transparency: contesting the logics of secrecy within the AI Act. European Law Open. 2023;2(1):79-105. doi:10.1017/elo.2022.47

European Commission (COM2021): The Artifial Intelligence Act, Proposal (2021). https://artificialintelligenceact.eu/wp-content/uploads/2022/05/AIA-COM-Proposal-21-April-21.pdf

European Union (PA2024): Artifial Intelligence Act. Text of the Provisional Agreement.https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf

European Union (AIACT2024): Artificial Intelligence Act (Regulation (EU) 2024/1689), Official Journal version of 13 June 2024. http://data.europa.eu/eli/reg/2024/1689/oj

Daneshjou R, Smith MP, Sun MD, Rotemberg V, Zou J. Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review. JAMA Dermatol. 2021 Nov 1;157(11):1362-1369. doi: 10.1001/jamadermatol.2021.3129. PMID: 34550305; PMCID: PMC9379852.

Gao, J., & Wang, D. (2023). Quantifying the Benefit of Artificial Intelligence for Scientific Research. ArXiv, abs/2304.10578.

Greenberg, B. (2016). 'Rethinking Technology Neutrality' (2016). 100 Minnesota Law Review (2016) 1495, pp. 1512–1513.

Guerrero O. & Margetts, H. (2024). Are All Policymakers Data Scientists Now? Data, Data Science and Evidence in Policymaking. LSE Public Policy Review. 2024; 3(3): 9, pp. 1–10. DOI: https://doi.org/10.31389/ lseppr.116

Haataja, M. & Bryson, J. (2023). The European Parliament´s AI Regulation: Should We Call It Progress? Amicus Curiae, Series 2, Vol 4, No 3, 707-718 (2023)

Ilyas, A., Santurkar, S., Tsipras, D., Engstrom, L., Tran, B. and Madry, A., (2019). Adversarials are Features, not Bugs. arXiv preprint arXiv:1905.02175.

Macnaghten, P., Davies, S. & Kearnes, M. (2015): Understanding Public Responses to Emerging Technologies: A Narrative Approach, Journal of Environmental Policy & Planning

Mele, C., & Russo-Spena, T. (2019). Innovation in sociomaterial practices: The case of IoE in the healthcare ecosystem. In Handbook of service science (pp. 517–544). Cham: Springer.

C. Mele, T. Russo Spena, V. Kaartemo, M.L. Marzullo (2021). Smart nudging: How cognitive technologies enable choice architectures for value co-creation. J. Bus. Res., 129 (2021), pp. 949-960, 10.1016/j.jbusres.2020.09.004

Mokyr, J., Vickers, C. & Ziebarth, N. (2015). "The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?" Journal of Economic Perspectives, 29 (3): 31-50.

Mylly, U-M. (2023). Transparent AI? Navigating Between Rules on Trade Secrets and Access to Information. IIC 54, 1013–1043 (2023). https://doi.org/10.1007/s40319-023-01328-5

Nersessian, D. & Mancha, R. (2020). From Automation to Autonomy: Legal and Ethical Responsibility Gaps in Artificial Intelligence Innovation, 27 Mich. Tech. L. Rev. 55 (2020). Available at: https://repository.law.umich.edu/mtlr/vol27/iss1/3

Nussbaum, M. C. 2006. Frontiers of Justice: Disability, Nationality, Species Membership. Cambridge, MA: Belknap Press.

Novelli, C., Casolari, F., Rotolo, A. et al. (2024). AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act. DISO 3, 13 (2024). https://doi.org/10.1007/s44206-024-00095-1

OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en.

Ott, T, & Dabrock P. Transparent human - (non-) transparent technology? The Janus-faced call for transparency in AI-based health care technologies. Front Genet. 2022 Aug 22;13:902960. doi: 10.3389/fgene.2022.902960. PMID: 36072654; PMCID: PMC9444183.

Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. Journal of Responsible Technology, 4 (2020), Article 100005, 10.1016/j.jrt.2020.100005

Robbins, S. A. (2019). Misdirected principle with a switch: explicit for AI. Minds & machines 29, 495-514.

Ruschemeier, H. & Hondrich, L. (2024). Automation bias in public administration – an interdisciplinary perspective from law and psychology, *Government Information Quarterly*, Volume 41, Issue 3, 2024. https://doi.org/10.1016/j.giq.2024.101953.

Shadikhodjaev, S. (2021). Technological Neutrality and Regulation of Digital Trade: How Far Can We Go?, European Journal of International Law, Volume 32, Issue 4, November 2021, pp. 1221–1247, https://doi.org/10.1093/ejil/chab054

Shue, H. 1996. Basic Rights: Subsistence, Affluence, and US Foreign Policy. Princeton, NJ: Princeton University Press.

Slattery, P., Saeri, A. K., Grundy, E. A. C., Graham, J., Noetel, M., Uuk, R., Dao, J., Pour, S., Casper, S., & Thompson, N. (2024). A systematic evidence review and common frame of reference for the risks from artificial intelligence. https://doi.org/10.48550/arXiv.2408.12622

Vaassen, B. AI, Opacity, and Personal Autonomy. Philos. Technol. 35, 88 (2022). https://doi.org/10.1007/s13347-022-00577-5

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual Explanations without ageing the Black box: automated decisions and the GDPR. Harvard Journal of law & technology, 31(2), 842-887.

Walmsley, J. (2020). Artificial intelligence and the value of transparency. AI and Society, 1–11.

M. Weinmann, C. Schneider, & J. vom Brocke (2016). Digital nudging. Business & Information Systems Engineering, 58 (6) (2016), pp. 433-436, 10.2139/ssrn.2708250

Wenker, K. (2022).A systematic literature review on persuasive technology at the workplace, Patterns,Volume 3, Issue 8, 2022, 100545, ISSN 2666-3899.

Wibye, J.V. (2022), Reviving the Distinction between Positive and Negative Human Rights. Ratio Juris, 35: 363-382. https://doi.org/10.1111/raju.12363

Viljanen, M. (2023). Menikö juna jo? Tekoälyn sääntelemisen mahdollisuuksista. Lakimies, 121(7-8), 1204–1231. https://journal.fi/lakimies/article/view/136265