Regulatory divergences in the draft AI act

Differences in public and private sector obligations

STUDY

Panel for the Future of Science and Technology

EPRS | European Parliamentary Research Service

Scientific Foresight Unit (STOA)
PE 729.507 – May 2022

EN

Regulatory divergences in the draft AI act

Differences in public and private sector obligations

This study identifies and examines sources of regulatory divergence within the AI act regarding the obligations and limitations upon public and private sector actors when using certain AI systems. A reflection upon possible impacts and consequences is provided, and a range of policy options is suggested for the European Parliament that could respond to the identified sources of divergence.

The study is specifically focused on three application areas of AI: manipulative AI, social scoring and biometric AI systems. Questions regarding how and when those systems are designated as prohibited or high-risk and the potentially diverging obligations towards public versus private sector actors and the rationale behind it, are described.

AUTHORS

This study has been 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) of the Secretariat of the European Parliament.

ADMINISTRATOR RESPONSIBLE

Philip BOUCHER, Scientific Foresight Unit (STOA)

To contact the publisher, please e-mail stoa@ep.europa.eu

LINGUISTIC VERSION

Original: EN

Manuscript completed in March 2022.

PE 729.507
ISBN: 978-92-846-9459-4
doi:10.2861/69586
QA-07-22-331-EN-N

http://www.europarl.europa.eu/stoa (STOA website)
http://www.eprs.ep.parl.union.eu (intranet)
http://www.europarl.europa.eu/thinktank (internet)
http://epthinktank.eu (blog)

Executive summary

This study **identifies and examines sources of regulatory divergence** within the AI act (AIA) regarding the obligations and limitations upon **public and private sector actors** when using certain AI systems. A reflection upon **possible impacts and consequences** is provided, and a range of policy options is suggested for the European Parliament that could respond to the identified sources of divergence. The study is specifically focused on three application areas of AI, being manipulative AI, social scoring and biometric AI systems. Questions regarding how and when those systems are designated as prohibited or high-risk and the potentially diverging obligations towards public vs. private sector actors and the rationale behind it, are described.

Through the use of existing examples of these three application areas both by public-and private actors, the potentially **diverging obligations** under the AIA are contextualised. These divergences in the AIA combined with current examples of AI applications in the three areas of analysis (biometrics, manipulative systems and social scoring), lead to an analysis per section. These three analyses will form the basis for discussion of key findings of the report, out of which follow a number of policy options.

Risks of AI systems

While not all AI systems harbour a potential harm for individuals, there are various examples of both public and private sector use of AI systems that have caused direct or indirect harms. The question is whether and how the AIA can mitigate risks caused by public or private sector AI, and how it overlaps or intertwines with other sources of EU law. This study finds that there is a **convergence of risks between public and private sector AI use**. As boundaries between service providers and users are blurring, and AI is increasingly part of a 'system of systems', conducting clear-cut risk assessments of AI systems will become increasingly challenging. In addition, the study documents tensions between general and sectoral regulatory approaches where appropriate. The AIA proposes procedural steps towards self-regulation of AI, much in line with the setup of the General Data Protection Regulation (GDPR), while at the same time proposing substantial measures in the form of, for instance, a prohibited list of AI applications. The governance of this list and the classification of high-risk applications or systems can lead to diverging interpretations and developments of AI systems. Moreover, the specific risk assessment proposed in the AIA can lead to diverging risk classifications in relation to, for instance, a risk assessment that needs to be performed on data as demanded by a GDPR for that same AI system. Regulatory coherence could be achieved by better aligning risk assessment efforts of digital (and AI-based) systems.

Prohibited practices and high-risk AI systems

The AIA exhibits a number of divergences in how it creates obligations for public and private actors. The study documents them and points to (challenges concerning) regulatory coherence with Union law where appropriate. We caution, however, that given the scope of this report we address only such divergences that directly or indirectly pertain to the broader discussion on public versus private obligations in relation to AI systems. Below, we summarise these diverging obligations as stipulated under the AIA's prohibited practices and high-risk systems.

Manipulative AI systems

We find that the tools for law enforcement to detect deep-fakes are considered to be high-risk, while deep-fakes themselves fall in the low-risk category. This is a peculiar divergence that appears to be grounded in the assumption that deep fakes (employed mostly by private actors for the time being) harbour less risks than AI systems in the hands of a public actor for the purpose of detecting deep fakes.

Regarding the AIA's regulatory (in)coherence with EU law on how it addresses manipulative practices, we find the Unfair Commercial Practices Directive (UCPD) to be most relevant. Thus, the divergence here is the scope of the ban on the use of manipulative AI systems for public actors versus the prohibition scope for private actors under the UCPD. In contrast to Article 5(2) UCPD, which caters to the protection of vulnerable groups beyond those strictly enumerated in Article 5(3) UCPD, the prohibition of certain manipulative systems as described in Article 5(b) AIA focuses only on vulnerabilities due to age and physical or mental disability. By not providing an analogous alternative, the AIA portrays a significant gap in the protection of persons who might be subject to AI manipulation on the basis of other protected characteristics under EU equality law, such as ethnicity, religion, race, sex, etc. Further, the AIA requires intent in order for the prohibition to be applicable, while its counterpart-article in the UCPD (Article 5(3) UCPD) **protects the defined vulnerable groups from commercial practices that are also unintentionally directed towards them**. A last AIA-regulatory incoherence with other EU law here is the narrow scope of the definition of harm in the AIA (physical or psychological harm), which **cannot be found elsewhere in Union law**. EU law usually speaks of harm in a generic way, without elaborating on the harm types that fall under it. The narrow scope of harm is as such divergent from general Union law.

Social scoring

The divergences in public versus private sector obligations in this type of AI applications relate to the fact that 1) the ban on social scoring for public authorities does not extend to the private sector; and 2) from a data-perspective, the grounds for prohibition for the public sector are more detailed (including categories on the basis of which data cannot be derived) than those on use for private actors – the latter are merely obliged to look at data quality and data governance.

The latter point brings us to the **AIA's regulatory incoherence in relation to the EU's data protection regime**. Externally, the use of social scoring by private actors creates issues between Article 10 AIA, Articles 22 and 35 GDPR and Articles 6 and 7 of the draft ePrivacy regulation. It is unclear how Article 10 AIA interacts and reconciles with data protection rights regarding consent and the right not to be subjected to automated decision-making and profiling, to name just a few.

Biometrics

The AIA **singles out law enforcement activities in publicly accessible spaces that employ real-time biometric identification systems (BIS)**, leaving out the use of real-time BIS by private sector actors. The AIA's approach to the ban in Article 5(1)(d) **differs from that in the GDPR, which does not distinguish between public and private data controllers**. Further, the prohibition of Article 5(1)(d) AIA focuses on BIS in "publicly accessible spaces" and appears to be **in direct contradiction to Recital (6) AIA**, which when clarifying the notion of AI systems explicitly refers to their effects "[…] in a physical or digital dimension'.' Lastly, while the deployment of biometric categorisation systems (BCS) and emotion recognition systems (ERS) as low-risk AI systems entrails mere transparency obligations, they do not apply to law enforcement. Regarding systems that can be just as intrusive as BIS (and are also prohibited for law enforcement), law enforcement here does not even have information or disclosure obligations.

Key findings

When evaluating our findings, we summarise the divergences and generalise their meaning for the AIA's purpose and normative outlook.

AIA divergences- treating similar AI systems differently depending on the user

Reflecting on the identified divergences, we notice that these relate to **treating similar practices (uses of AI systems) differently depending on the actors that deploy them**. We see the latter in the dichotomy of public versus private sector obligations in relation to social scoring, as well as in relation to the prohibition of real-time BIS for law enforcement. Our analysis shows, however, that

the separation of private and public actors' AI practices is less and less defendable, as the **risk levels associated with AI use by either public or private actors do not differ in the power asymmetry they create towards the individual**.

Further, the example of designating systems that law enforcement uses for the detection of deep-fakes as high-risk versus designating deep-fakes low risk in general, provides additional evidence for the lacking in the AIA's risk-based approach. We see **similar types of systems placed in a different risk category** without backing up the rationale to do so with concrete risk level assessment criteria.

Regulatory coherence with EU law

Certain scoping issues relate to the requirements of harm and intentionality and their relationship to the UCPD and other sources of Union law, as well as to the distinction of public versus private upheld by the AIA in contrast to the GDPR. The **scoping issues portray an urgent need for more harmonisation of the AIA provisions with existing EU law**.

The procedural lack of coherence brings more obvious inconsistencies between the AIA on the one hand, and the GDPR and the ePrivacy regulation on the other. One key example would be that the **proper procedures around training data for AI systems including the obtaining of consent and from whom, as well as the legal basis for such processing, are unclear**.

Policy options

Address the incoherence of risk assessment and introduce explicit risk criteria: The AIA is different from other recent EU legislative endeavours such as the GDPR in the sense that the normal risk-based approach to regulating a technology has been expanded from being a procedural obligation to also add a substantial part. The act proposes a classification of risks in relation to AI by dividing applications of AI in three categories: prohibited, high-risk and low-risk. However, as evidenced by the identified divergences, the AIA's risk categories are not always applied consistently when it comes to public or private actors and their obligations to mitigate such risks. When looking at risk assessment, a policy option would be to make very clear what the risk assessment is precisely about and to provide clear delineations or cut-off points on what part of the system needs assessment. In addition, providing guidelines on how the AIA risk assessment interacts with other risk assessment obligations put forward in many of the EU regulations and directives that deal with digitisation (e.g. the data protection impact assessment (DPIA) in the GDPR) would be a step towards harmonisation.

Consider strengthening information and disclosure obligations with withdrawal rights: Transparency in the AIA is not linked to a subjective right and remains as such at the level of principle or policy aspiration. Further, as a disclosure obligation it is not applicable to law enforcement, limiting thereby even further chances for individual protection. An option to overcome this unsatisfactory state within the AIA would be 1) to clarify and directly stipulate in the AIA's provisions how GDPR rights and remedies are applicable to the addressees of AI systems, especially so when data rights are involved; and 2) to further critically assess the connection between the AIA's transparency obligations and redress mechanisms by strengthening information and disclosure obligations with withdrawal rights.

Consider non-linear modes of governing and co-regulation strategies: The AIA's current approach of governing is hierarchical (law-centric) combined with forms of self-regulation (techno-centric). Obligations that are not clarified in a top-down manner are left to the industry standardisation bodies to figure out, which severs the channels of communication. More importantly, this approach focuses largely on the material features of AI systems, omitting to incorporate in a more consistent way the underlying or emerging socio-technical changes and their actual impacts on individuals and society. The AIA does not provide clear measures in place to

monitor such regulatory effects, except the ex-ante risk assessment and perhaps 'by-design' approaches via regulatory sandboxes. Measuring long-term effects and socio-technical changes as a result of using AI system by means of ex-post impact assessments is currently lacking. The AIA can be re-evaluated in this fashion to consider 1) **the trajectory and distribution of AI systems**, including the factors that drive its proliferation in sectors; 2) **the political viability of the regulation** and how certain features affects stakeholder perceptions; and 3) **potential ways for regulatory leverage**. 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.

Table of contents

List of abbreviations

AIA	Artificial intelligence act
BCS	Biometric Categorisation Systems
BIS	Biometric Identification Systems
DA	Data Act
DMA	Digital markets act
DPIA	Data Protection Impact Assessment
DSA	Digital services act
DGA	Data Governance Act
EC	European Commission
ERS	Emotion Recognition Systems
EU	European Union
GDPR	General Data Protection Regulation
LED	Law Enforcement Directive
UCPD	Unfair Commercial Practices Directive

1. Background

Artificial intelligence's (AI) uptake and use across all sectors of society is increasingly seen as the determinant of a great-power status in matters both political and economic[1]. The latest addition to the landscape of the digital age is subject to fierce international competition, all the while **regulators puzzle on how to vest AI developments in a meaningful legal framework** that **protects citizens' rights, boosts digital sovereignty and creates enough legal certainty** for all stakeholders involved in the AI chain. The latter is needed not only for those who develop and deploy AI systems and whose processes would benefit from streamlined rules of the road, but also for those who are the direct addressees of (unaccounted for) algorithmic harms.

The proposal for a Regulation concerning AI (the AI Act or AIA) presented by the European Commission (EC) on 21 April 2021[2] is one of the first attempts at horizontal AI regulation[3] that harmonises rules for the development, placement on the market and use of AI systems. Following in the footsteps of the General Data Protection Regulation (GDPR), by means of which the EU propagated norms on data regulation beyond its borders, the draft AIA is a similar attempt to gain a head start in AI governance. The latter is facilitated by its *aterritorial*[4] rationale – the Act would extend the EU's jurisdiction to all AI systems that produce outcomes within the EU, irrespective of whether the system's user or provider are located within the EU. Further, the AIA **proposes different regulatory burdens depending on the AI system** at hand – it *bans* certain types of systems, *regulates* under the umbrella of a high-risk regime such that pose a threat to fundamental rights and safety, and places *voluntary* constraints on less risky systems –, thereby following a risk-based approach.

This report focuses on regulatory choices within the Act that create **diverging obligations for public and private actors around prohibited and high-risk AI** systems such as manipulative AI, social scoring and biometrics, and the rationale behind them. For instance, Art. 5(1)(c) AIA prohibits the use of social scoring systems by public authorities, **without banning such systems deployed by private actors**. The latter systems, however, oversee financial flows, insurance policies and claims, housing applications, etc. and **control as such access to essential (state-like) services**, which individuals cannot forego. Thus, while opting out is not a viable option, the scoring systems used by private providers of such services are **subject to less stricter standards** than the ones for services of similar magnitude in the public sector. Another example is found in the prohibition of Art. 5(1)(d) AIA, which covers biometric facial recognition by law enforcement, while leaving the practice open for other public authorities and the private sector. Private actors, especially, are increasingly making use of image recognition cameras in combination with biometric technology to provide access to shops, banks, etc.[5], or to instantaneously assess a person's reaction to a product or situation, and to thereby attain a better position in inducing desired consumer behaviour. Next to the more obvious implications of **indiscriminate (consumer) surveillance** or discrimination

---

1 European Parliament/Special Committee on Artificial Intelligence in a Digital Age, Draft report on artificial intelligence in a digital age (2020/2266(INI), PR_INI (europa.eu), p. 8.

2 European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021) 206 final).

3 Another noteworthy attempt is, for instance, the US National AI Initiative Act, which came into force on 1 January 2021.

4 Floridi, L., 'The European Legislation on AI: a Brief Analysis of its Philosophical Approach.' *Philosophy & Technology* (2021), pp. 1-8.

5 See: AI Facial Recognition and IP Surveillance for Smart Retail, Banking, and the Enterprise, Interesting engineering.

against certain population groups towards which the algorithm might be less sensitive, the use of biometric identification systems in the last example **structurally affects the way we engage with daily occurrences or give consent**.

The report places in context these and other examples of **diverging obligations** under the AI Act by relating them to applications of AI systems and their harms in practice, as well as to other instances of Union law. By means of the comparison between the AIA and other EU regulatory instruments, as well as between the outcomes/harms of AI systems deployed by public vs. private actors, the study seeks to establish **to what extent the AIA adequately addresses the responsibilities that come along with the deployment of societally-transformative technologies**. In the same spirit, the report also questions traditional, linear conceptions of agency and role division, highlighting the **AI industry's 'private ordering'**[6] – the regulatory power digital technologies have over us – and its regulatory effects, and bringing them to the attention of legislators, regulators and policy-makers in ongoing discussions.

While we investigate regulatory divergence relating to prohibited and high-risk AI systems in the AIA, we do so through the conceptual lens of legislative *coherence*. As it will be explained in detail below, we employ coherence to assess whether the AIA's design and provisions do justice to its intention and normative outlook, as well as to other principles of law on related topics found in existing and upcoming legal instruments related to or applicable to Europe's digital agenda, of which the AI Act is part. We thus not only examine **the Act's suitability for effectively regulating** biometrics, face recognition and/or social scoring AI systems, but also identify the implications of the established discrepancies in public vs. private ordering in the strive for legal certainly and harmonisation of digital regulations.

This report proceeds as follows. Chapter 2 describes the research scope, employed concepts and methodology. Further, since the AIA is (at least partially) a risk-based instrument and regulates AI systems according to risk levels, to better substantiate our argument chapter 3 elaborates on conceptions of risk and AI harms, among others related to the use of AI systems by public and private actors. To pave the way into the AIA's understanding, chapter 4 starts with an overview of the AIA's stated objective, as well as of the objective of the regulatory package surrounding the EC's digital strategy. Chapter 5 continues by diving into the selected AI systems – manipulative AI systems, social scoring and biometric systems. We provide an overview of the systems' underlying technology, before elaborating on the provisions of the AIA that address such systems and their use in both the public and the private sector. Chapter 6 summarises the identified divergences before offering policy options in Chapter 7.

[6] Delacroix, S., 'Beware of 'Algorithmic Regulation'', *SSRN*, 2019.

2. Research scope, concepts and methodology

2.1. Scope

This project aims **to identify and examine sources of regulatory divergence** within the AIA regarding how **public and private sector actors may use certain AI systems**, to reflect upon their **possible impacts and consequences**, and to develop and assess a range of policy options for the European Parliament that could respond to the identified issues. The latter requires consulting a number of existing and proposed sources of EU legislation, which are either context- or rationale-fitting to the research questions at hand, and screening them within the short time frame of the study. Therefore, **we scope our research as follows**:

1) We examine **in particular the Recital, and Titles II, III and IV of the AI Act**, dedicated to prohibited AI practices, to creating rules for AI systems with high-risk AI to the health and safety or fundamental rights, and to transparency obligations for certain AI systems respectively.

2) As the AI Act is part of the regulatory package accompanying the Commission's strategy 'Europe fit for the digital age', though not in depth, **we consider some of the regulatory instruments currently under negotiation**:

   a) the draft Digital Services Act with provisions on recommenders and research data access[7];
   b) the draft Data Governance Act (DGA), concerning data sharing frameworks[8];
   c) the draft ePrivacy Regulation (updated rules for all electronic communications including cookies, spam, consent, content and metadata)[9].

3) Lastly, we **consult existing EU legislation**, whose **regulatory rationale has impacted the structure**[10] **and unfolding** of the AI Act, or interplays with the Act's proposed provisions:

   a) EU product safety law[11];
   b) the Unfair Commercial Practices Directive (UCPD)[12];
   c) the Law Enforcement Directive (LED)[13];

---

7 European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM(2020) 825 final).

8 European Commission, Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) (COM(2020) 767 final).

9 European Commission, Proposal for a Regulation on Privacy and Electronic Communications, European Commission website.

10 See Veale, M., and Zuiderveen Borgesius. F., 'Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach.' *Computer Law Review International 22(.4)*, 2021, pp. 97-112.

11 Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC, OJ L 218/82.

12 Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business to consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), OJ L 149/22, art 5.

13 Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119/89 ('Law Enforcement Directive').

d) the General Data Protection Regulation (GDPR)[14].

When addressing diverging obligations within the AI Act itself we refer to them as **'AIA-divergences'**, while for such between AIA provisions and other EU legal instruments we speak of **regulatory (in)coherence** or harmonisation. In the following, we explain the method of our study.

2.2. Concepts and methodology

In this step of the analysis we determine the conceptual lens for our inquiry into the legislative framework outlined in the previous section, and introduce the practical research steps that instrumentalise and contribute to the application of the conceptual lens.

Regulatory coherence stems from doing justice to the regulation's **underlying rationale and normative outlook** on the one hand[15], and from **effectively supporting that normative outlook** and established principles of law on the same topic on the other hand[16]. The so portrayed two-dimensional concept of coherence is **well-suited for the evaluation of provisions within a single regulation**, as well as for **evaluating a single piece of legislation within a broader regulatory regime**. Upon identifying a number of divergences, we use coherence as a lens to evaluate how any such divergences or inaccuracies in regulating prohibited or high-risk AI systems **affect the AI Act's objective**, as well as the objective of the broader regulatory framework surrounding the EU's digital agenda.

Practically, we conduct a **theoretical examination** of the regulation's elements and composition. We do so in a twofold manner. We first use **a 'grounded method'** to scrutinize the texts for various kinds of issues addressed and the mobilised arguments, and collect the information pertaining to the categories **'type of risk'** (what kind of AI harm is being averted), **'target audience'** (public, private or both), **'type of obligation'** (prohibition, exceptions, requirements). Further, next to examining the laws' content, we use the **'law-in-context' research methodology** as conceptualized by Sacco and further systematized by Van Hoecke[17]. This method allows us to ask questions about the social reality as well, thereby taking into account the historical and socio-economic context of the laws under investigation. This approach fits well with our purposes and research design, as we consider not only upcoming but also existing EU legislation within a new social reality and domains conditioned by the digital era and contemporary AI development. Our method is thus **partially descriptive and partially analytical**.

Once we examine the legal texts in the identified categories and their rationales, we compare how similar AI-related risks have been dealt with regarding public or private use of AI systems and why, and record our findings in a draft report. In the process of doing so, we also draw insights from examples on the use of AI systems by public and private parties, and what focal points those reveal. We then provide key findings and policy options.

---

[14] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119/1.

[15] See on the latter Koops, Bert-Jaap. 'Ten Dimensions of Technology Regulation-Finding Your Bearings in the Research Space of an Emerging Discipline.' *Dimensions of technology regulation*, 2010, p. 319.

[16] Guihot, M. 'Coherence in technology law.' *Law, Innovation and Technology*, 11(2), 2019, pp. 311-342.

[17] Van Hoecke, M. and Warrington, M., 'Legal cultures, legal paradigms and legal doctrine: towards a new model for comparative law.' *International & Comparative Law Quarterly*, 47(3), 1998, pp. 495-536.

3. Risks of AI systems

While not all AI systems harbour a potential harm for individuals[18], we have witnessed both public and private sector use of AI systems that has caused direct or indirect harms[19]. This raises questions on **what harm is or can be in relation to AI systems, and on how it manifests and relates to conceptions of risk**. The continuous collection of examples of harms caused by AI systems, as done by NGOs such as AlgorithmWatch[20] and Amnesty International[21] are crucial to better understand AI's direct and indirect impact on society, and to inform and keep score of actual harms manifested through AI systems.

Harm in a legal sense refers to damages or loss to a person or group of persons (be they natural or legal persons)[22]. Further, there are intricate socio-legal and socio-political inroads to explaining our current understanding and use of the notion of risk in relation to harms, and the relation between risk and rights. While a comprehensive analyses thereof is beyond the scope of this report, here it suffices to state that **risk is often explained as the chance of a particular harm to occur**:

> *Risk is […] a technique for creating knowledge and certainty about future events that are uncertain […] (Bernstein, 1998). It is seen as the chance ([…] likelihood) that a danger (i.e., an event with harmful consequences) will happen. Risk assessment methodologies are built upon the assumption of forecasting the future on the basis of statistics and probabilities*[23].

Though harm in relation to AI can be approached from different perspectives – including technological (mainly in the AI-subdomain of machine learning (ML)), ethical (dealing with the moral underpinnings of using/not using specific AI systems to tackle a challenge), and through the viewpoint of the citizens (how they interpret AI developments, what are their hopes or fears about it) – for the sake of this study we adopt a legal perspective. The latter deals with questions on **how parties can be held accountable for harms**. Other related questions hereto examine whether current legal frameworks and systems of checks and balances are equipped to deal with AI risks and harms.

3.1. Digitisation, risks and harms

The potential **harms of public sector AI** as pointed out by courts or by public opinion[24] evolve among others around **information asymmetry, not knowing about automated decision-making processes in government services, not being aware of and not being informed about being part of a dataset** used in such application, or being part of the outcome of a 'fact-finding' algorithm that predicts a certain likelihood of a citizen falling into a particular category. Especially in policy areas such as policing and fraud detection, once wrongly on such a list, it can be life-destroying (a 'red-flag' may pop up in requests for benefits, childcare services, job-seeking etc.). Yet, where perhaps new checks and balances or new interpretations of existing checks and balances are necessary within the public sector itself when it comes to the use of AI systems in the public sector,

---

18 For example, advanced algorithms which use data about tomatoes would not fall in this category.

19 SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens, rechtspraak.nl.

20 New report highlights the risks of AI on fundamental rights, Algorithm Watch.

21 Ban dangerous facial recognition technology that amplifies racist policing, Amnesty International.

22 Kleinig, J., 'Crime and the Concept of Harm.' *American Philosophical Quarterly*, 15(1), 1978, pp. 27-36.

23 Van Dijk, N., Gellert, R., & Rommetveit, K., 'A risk to a right? Beyond data protection risk assessments.' *Computer Law & Security Review*, 32(2), 2016, pp. 286-306.

24 BEUC., Artificial Intelligence what consumers say, 2020.

further risks lie in **the role of the private sector in co-developing public services**, and the **risks posed by consumer-driven 'big tech' platforms** that converge with public services.

Recent uses of deep-fakes and election meddling by *Facebook* (now *META*) spinoff companies[25] have shown that big tech has little consideration for human rights or upholding democratic values[26]. However, despite many fines imposed by the EU in a semi-constant stream towards private actors breaching GDPR[27], there is little evidence for change in behaviour/ logic within digital platforms, and it is such **platforms that currently drive AI development**. This lack of (self) regulation among private digital 'horizontal' actors is partially due to the logics of data platforms; as they do not belong to a particular sector, there is no or little sectorial legislation or code of conduct in place[28]. Another reason can be found in what data has done in (and to) society: it has **converged consumers with citizens and companies with governments**. Boundaries between service provider and user are blurring – we are seeing heads of state discussing policy on social media and consumer platforms taking over communicating channels for public services, for example. This makes risk assessment of an AI system very challenging.

Many harms attributed to AI stem from this convergence and from the logics of 'disruption' we are already familiar with from the era of Big Data. Yet, **some new harms have built on top of data-related harms as a result of AI**. Scholars have mentioned computational and automated violations of privacy, behaviour influencing through hyper-personalisation, algorithmic opacity (not knowing if or when you are dealing with an AI system, and what predictions it is making about you), lack of diversity of norms that get built in and automatically enforced[29], dehumanisation through hard-coding human interaction, irreproducibility of AI outcomes which can lead to monopolies of being able to deal with complex AI systems, the negation of novel and public futures because data-driven AI systems are trained on things that happened in the past[30], and finally power asymmetries enhanced by algorithms without clear possibilities for redress[31]. Specific forms of AI, most notably ML and the application thereof in decision-making processes, be they in the public or private domain, pose novel risks that go beyond the somewhat known risks and harms of 'datafying' society[32]. The AIA addresses physical or financial harms, to a certain extent 'on top of' existing regulatory frameworks that would be in pace to address other forms of harm (such as psychological or immaterial forms of harm). The question is whether and how the AIA can mitigate risks caused by public-sector or private-sector AI systems and how it overlaps or intertwines with other EU directives or regulations.

---

25 Cambridge Analytica and Facebook: The Scandal and the Fallout So Far, The New York Times, April 2018.

26 A case in point here is the recent legal action of Rohingya people against Facebook for hate speech. See: Rohingya refugees sue Facebook for $150 billion over Myanmar violence, Reuters, December 2021.

27 EU targets Big Tech with 'hit list' facing tougher rules, Financial Times.

28 They often position themselves merely a messenger that helps other companies through advertising and targeted marketing The EU has recently proposed policy on platform liability to try and restore some of the power imbalance, through the Digital Services Act, see The Digital Services Act package, European Commission Website.

29 The Markup, Algorithms Behaving Badly: 2020 Edition, December 2020.

30 Opinion: Billionaire capitalists are designing humanity's future. Don't let them, The Guardian, February 2021.

31 See Human-centred AI in the EU, YouTube for more elaboration on risks and potential harms.

32 7 Types of AI Risk and How to Mitigate their Impact, Towards Data Science, September 2020.

3.2. How to regulate as a result of risk: Sector-specific rules for private actors and generic rules for the public sector?

The European Parliament has come to the conclusion that as with any novel technology some basic legal questions need to be taken into account before deciding on a regulation strategy (emphasis added)[33]:

> *(i)* ***Which rules apply in this sector, and what are the rationales for those rules?*** *A rule may, for example, aim to protect a human right, or express a legal principle, such as equality, contractual freedom, or the right to a fair trial. Economic rationales also differ from sector to sector. […]*
>
> *(ii)* ***How is or could AI decision-making be used in this sector****, and what are the risks? For instance, false positives are a serious problem in the context of criminal law […]. By contrast: if an incorrect decision by an AI system for price discrimination makes a consumer pay extra, the effect is often less harmful than when an incorrect AI decision leads to someone being arrested by the police.*
>
> *(iii) Considering the rationales for the rules in this sector,* ***should the law be improved in the light of AI decision-making****? Does AI threaten the law's underlying principles or undermine the law's goals? If current law leaves important risks unaddressed, amendments should be considered […][34].*

A recent report by the standardisation body CEN-CENELEC address the need to **rethink liability in the case of AI**, because the harmed party can never manage to access and produce proof of all links in a digital network, let alone understand what happened in the AI model that led to certain choices, actions and consequences in the real world. As AI systems will inevitably complicate things, the regulator should have a grasp of the complexity, while '*stakeholders need to understand whether the regulation applies to an organization, product or service. Any legislation should be formulated around the specific properties of the system (e.g. ML or statistical inference) and application context*'[35]. This last advice, stemming from public consultation responses on the EU AI Whitepaper, **poses a direct opposite to the *lex generalis* approach of the AI Act**. It further makes explicit the tensions between general and sectoral regulation approaches, and between self-regulation and the development of sector-specific standards vis-a-vis a more general and human-rights based scheme to protect citizens and consumers of harms caused by AI systems. Another crucial element in the development of AI regulation is to keep in mind who it is for: the aim is to make sure society at large benefits from the technology without harming individuals or groups.

---

33 AI rules: what the European Parliament wants, European Parliament website.

34 CEN-CENELEC, CEN-CLC Response to EC White Paper on AI, June 2020.

35 CEN-CENELEC, CEN-CLC Response to EC White Paper on AI, June 2020.

4. The AI Act in context

4.1. The AI Act's objective

The EC's proposed AI Act aims to give developers, deployers and users of AI systems clear requirements and obligations regarding specific uses of AI, while at the same time also reducing administrative and financial burdens for businesses, in particular SMEs[36]. Protection of EU values, ethical principles and fundamental rights is said to play an important role in the Act, in line with the EC's vision of AI as a trustworthy tool for people and a force for good in society that increases human well-being. Accordingly, the rules for AI in the Act are to cater to a human-centric rationale. At the same time, the Act supports the EU's strategic goal to be a global leader in the development of trustworthy and ethical AI. The AIA thus formulates following specific objectives:

> *'Ensure that AI systems place on the Union market and used are safe and respect existing law on fundamental rights and Union values;*
>
> *Ensure legal certainty to facilitate investment and innovation in AI;*
>
> *Enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;*
>
> *Facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.'*[37]

As described in the introduction, those objectives are approached by means of a risk-based approach, producing an overview of prohibited AI practices, high-risk systems, limited risk systems and minimal risk systems. Prohibited risk systems include those AI systems of which the use is regarded as unacceptable and contravening values of the EU, such as violating human and fundamental rights. The list of prohibited systems covers applications with significant potential of manipulation of people or exploitation of specific vulnerable groups like children or person with disabilities, AI-based social scoring for general purposes or the use of real time remote biometric identification systems in publicly accessible spaces by law enforcement purposes[38].

High-risk systems include AI technology used in critical infrastructures, educational or vocational training, safety components of products, employment, workers management and access to self-employment, essential private and public services, law enforcement that may interfere with people's fundamental rights, migration, asylum and border control management. For these high-risk systems the Act stipulates additional requirements, such as conformity assessments by a third party, risk management system, record keeping of events while the system is operating, technical documentation before the system's placement on the market, human oversight and transparency, and provision of information to users. For limited risk systems, e.g. chatbots, there are, for instance, transparency obligations aimed at making sure users are aware that they are interacting with a machine. Minimal risk systems are AI systems such as AI-enabled video games or spam filters[39]. The scope of the regulatory framework does not include the development and use of AI for military purposes.

---

36 Regulatory framework on AI, European Commission website.

37 European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021) 206 final).

38 Ibid.

39 Regulatory framework on AI, European Commission website.

4.2. The regulatory package accompanying the EC's strategy 'Europe fit for the digital age'

The AI Act is best analysed in context of the rest of the regulatory package accompanying the Commission's strategy 'Europe fit for the digital age.' This strategy aims at making the digital transformation work for people and businesses, while at the same time helping to achieve Europe's target of being climate-neutral by 2050. Part and parcel of the strategy are the EC's objective to strengthen Europe's digital sovereignty, to become a role model for the digital economy, to set standards on data, technology and infrastructure, and to support developing economies in their digitalisation process[40].

Europe's digital future is conceptualised in three pillars. The first pillar focuses on technology that works for people, part of which are the AI Act's objectives described in the previous section. The second pillar focuses on a fair and competitive digital economy. An important role in this is reserved for the proposed Digital Services Act which aims to strengthen the responsibility of online platforms and to clarify rules for online services. Another aim of this pillar is to increase access to high-quality data, while ensuring that personal and sensitive data is protected[41]. The proposed DGA and the GDPR contribute to this goal. The third pillar is an open, democratic and sustainable society, aiming, for instance, to empower citizens with better control and protection of their data, to fight disinformation online and to foster diverse and reliable media content[42].

Other regulatory instruments part of this regulatory package are currently under negotiation – the draft Digital Services Act (with provisions on recommenders and research data access)[43]; the draft Digital Markets Act (with provisions on AI-relevant hardware, operating systems and software distribution)[44]; the draft Machinery Regulation (revising the Machinery Directive in relation to AI, health and safety, and machinery)[45]; and the announced product liability revision relating to AI[46]. The Data Act[47] as the last piece of the puzzle, has just been published at the time of writing, and builds on the Free Flow of Data initiative set out by the former Commission with the goal to accelerate the sharing and usage of non-personal data in Europe.

The aim of the proposed Digital Services Act and proposed Digital Markets Act is to create a safer digital space, in which the fundamental rights of citizens are protected, and a level playing field for businesses is established[48]. With the Digital Services Act the Commission hopes to improve the mechanisms for the protection of fundamental rights of users and the removal of illegal content. In

---

40 See: A Europe fit for the digital age, European Commission website. And Shaping Europe's digital future, European Commission, February 2020.

41 See: Shaping Europe's digital future, European Commission, February 2020.

42 Ibid.

43 European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM(2020) 825 final).

44 European Commission, Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) (COM(2020) 842 final).

45 European Commission, Proposal for a Regulation of the European Parliament and of the Council on machinery products (COM(2021) 202 final) (Machinery Regulation).

46 See European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Fostering a European Approach to Artificial Intelligence (COM(3032) 205 Final)' (21 April 2021) p. 2.

47 Data Act: Commission proposes measures for a fair and innovative data economy, press release, European Commission, 23 February 2022.

48 The Digital Services Act package, European Commission website.

addition, the EC aims to create stronger oversight and transparency measures for online platforms, including rules on algorithms used for recommendations[49].

In February 2020 the Commission published a report on the safety and liability implications of AI, the Internet of Things and Robotics, addressing the emergence of new challenges in terms of product safety and liability such as autonomy, data dependency, connectivity, complexity of products and systems, etc. in relation to these technologies. The report highlights current product safety legislation gaps, in particular such in the General Product Safety Directive, Machinery Directive, the Radio-Equipment Directive and the New Legislative Framework[50]. Accordingly, a new draft Machinery Regulation aims to tackle a number of problems, including new risks originating from emerging technologies and insufficient provisions for high-risk machines[51]. The proposed Data Governance Act aims to stimulate the availability of data by strengthening data-sharing mechanisms and improving trust in data intermediaries across the EU[52]. The aim of the instrument is to make public sector data available for re-use, to stimulate data sharing between businesses and to allow personal data to be shared with the help of a personal data-sharing intermediary[53]. In the next section, we will delve into the divergence that we have encountered inside the AIA when it comes to consequences of the Act for public - and private actors.

49 The Digital Services Act: ensuring a safe and accountable online environment, European Commission website.

50 Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64, European Commission, 19 February 2020.

51 European Commission, Proposal for a Regulation of the European Parliament and of the Council on machinery products (COM(2021) 202 final) (Machinery Regulation).

52 See: European Data governance act, European Commission website.

53 European Commission, Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) (COM(2020) 767 final).

5. Public-Private divergences in the AI Act

This chapter examines the identified divergences between obligations for public and private sector actors. It looks at parts of the AIA that lay down rules for the use of manipulative AI systems, social scoring and different biometric identification systems in order to identify the obligations for public and private sector actors and corresponding divergences therein. We start with the **wording of the AI Act in relation to an application area**. Upon sketching the systems and **the rules the AIA attaches to their use**, we provide **examples of their deployment by public and private actors**, and discuss the implications of the systems' usage and **identified divergences** in regulating them.

5.1. Identified divergences

	Public sector	Private sector
Manipulative AI systems	Art. 5(1)(a),(b) AIA bans the use of manipulative AI systems by public actors in case of intentional harm and for a limited number of exploited vulnerabilities	Art. 5(2) UCPD protects against the commercial exploitation of additional vulnerabilities beyond the ones listed in Art. 5(3) UCPD Art. 5(3) UCPD protects vulnerable groups also 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 ban 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	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 wrt their use of ERS and BCS	Art. 52(2) AIA envisions transparency obligations for the use of ERS and BCS by private sector actors

The table above portrays the diverging obligations for public and private sector actors within the AIA that our analysis has identified. While the prohibition on the use of manipulative AI systems applies to both public and private actors, it is the prohibition's intent requirement as well as the limited number of vulnerabilities it protects against exploitation that play out differently in relation to other Union law, and portray an explicit regulatory divergence between public obligations under the AIA and private ones under the UCPD. A further divergence is identified in how law enforcement AI systems that detect deep fakes are handled (high-risk) vs. the labelling of deep-fakes themselves as low risk in the hands of private actors.

Further, with regard to social scoring we reflect on the ban of the practice for the public sector as opposed to designating social scoring AI systems used by the private sector high-risk. These diverging obligations manifest themselves in further data-related differences for public and private actors, which have also ramifications for the AIA's coherence with other Union law.

Lastly, we see a similar dichotomy in the obligations for public and private sector actors regarding biometric AI systems – a ban for law enforcement vs. a high-risk classification for the private sector deployment of real-rime BIS. Upon representing this divergence, we also reflect on the lacking transparency obligations for law enforcement in the use of ERS and BCS, which however are there for private actors.

Overall, the divergences question the risk-assessment criteria and approach of the AIA, as well as its alignment with fundamental rights and values.

5.2. Manipulative AI systems

5.2.1. AIA divergences

	Public sector	Private sector
Manipulative AI systems	Art. 5(1)(a),(b) AIA bans the use of manipulative AI systems by public actors in case of intentional harm and for a limited number of exploited vulnerabilities	Art. 5(2) UCPD protects against the commercial exploitation of additional vulnerabilities beyond the ones listed in Art. 5(3) UCPD Art. 5(3) UCPD protects vulnerable groups also 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 does not apply to law enforcement	Art. 52(3) AIA envisions only transparency obligations for deep fakes

Art. 5(1)(a) and (b) AIA prohibit:

> *(a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm;*
>
> *(b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm;*

Though the ban on manipulative AI systems applies equally to both public and private actors, **here it is the prohibition's intent requirement that portrays an explicit regulatory divergence between public obligations under the AIA and private ones under the UCPD**. The latter protects consumers against unfair commercial practices of businesses and addresses as such largely the private sector. The ban of the AI Act is dependent on establishing the deployer's intent ('in order to'). The latter means that even if the system is manipulative in itself and harm has manifested, the

prohibition would not be triggered unless the AI-facilitated manipulation has resulted in *intentional* harm. In contrast, the UCPD which protects consumers against unfair commercial practices and which the AIA complements[54], protects from commercial practices that are also *unintentionally* directed towards the groups protected in Art. 5(3) UCPD. This means that consumers who have experienced, for instance, financial harm as a consequence of the unfair (e.g. manipulative) commercial practices of private actors, have a better chance at legal protection than individuals who have been harmed by AI systems that misbehave during their deployment[55] by public actors.

Further, by focusing only on vulnerabilities due to age, physical or mental disability, Art. 5(1)(b) AIA mirrors quite exactly Art. 5(3) UCPD[56]. However, the UCPD provides protection to other vulnerable persons beyond the strictly enumerated ones through the subgroups identified in Art. 5(2) UCPD. The **AIA does not provide an analogous alternative**, and portrays as such a significant gap[57] in the protection of persons who might be subject to AI manipulation on the basis of other protected characteristics under EU equality law such as ethnicity, religion, race, sex, etc. Often, in well-documented disinformation examples, **cognitive or emotional manipulation is facilitated by the interplay of a number of these characteristics**. Here again the public ban of the AIA would be weaker than the one aimed at private actors under the UCPD.

Further divergences relate to the transparency obligation for public and private parties in relation to deep-fakes. The AIA considers deep-fakes a low risk technology, attaching to their overall use and deployment minimum requirements including such of transparency under Art. 52(3) AIA. This provision requires creators of deep-fakes to disclose to the receiving party that the content has been manipulated. However, according to Art. 52(3)2 AIA this labelling obligation does not apply to law enforcement. This means that when these particular public actors use deep-fakes, they do not need to disclose it.

In contrast, Annex III 6(c) AIA lists 'AI systems intended to be used by law enforcement […] to detect deep-fakes' as high-risk. This approach treats here similar AI systems differently - **deep fakes appear to carry less risk than the methods used by public actors to detect them**[58]. Detection systems are thus only allowed under the requirements of Chapter 2 and 3 AIA (the high-risk regime), which

---

54 See Recital 16 of the AI Act.

55 Hacker, P. Manipulation by algorithms. Exploring the triangle of unfair commercial practice, data protection and privacy law, *European Law Journal*, 2021, p. 28.

56 Art 5 UCPD reads:

1. Unfair commercial practices shall be prohibited.

2. A commercial practice shall be unfair if:

(a) it is contrary to the requirements of professional diligence,

and

(b) it materially distorts or is likely to materially distort the economic behaviour with regard to the product of the average consumer whom it reaches or to whom it is addressed, or of the average member of the group when a commercial practice is directed to a particular group of consumers.

*3. Commercial practices which are likely to materially distort the economic behaviour only of a clearly identifiable group of consumers who are particularly vulnerable to the practice or the underlying product because of **their mental or physical infirmity, age or credulity** in a way which the trader could reasonably be expected to foresee, shall be assessed from the perspective of the average member of that group. This is without prejudice to the common and legitimate advertising practice of making exaggerated statements or statements which are not meant to be taken literally. **(Emphasis added)***

[…]

57 Hacker, 2021, p. 28.

58 Mahler, T., 'Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal.' *Nordic Yearbook of Law and Informatics*, 2021.

mandate among others record keeping, risk-management systems and data governance frameworks to be in place.

5.2.2. Examples of use of deep-fakes from the practice

There is a growing number of companies offering technology that can create deep-fakes: fabricating audio and video representations of people saying or doing things that did not happen but are increasingly difficult to distinguish from real ones[59]. For example, Descript offers an application that uses a text-to-speech model, e.g. for editing recordings. By typing in text, text will be turned into speech, and the voice used can be one's own voice or such from a selection of 'stock voices'[60]. Similar technology was used in the deep-fake videos that went viral on Facebook, portraying Nancy Pelosi intoxicated while giving an official address[61]. Used as a revenge by supporters of her political opponents, social media users sought to undermine her political standing.

Deep-fakes have also found their way in political diplomacy[62]. In 2021, **members of parliament** in Latvia, UK, Estonia, Lithuania[63] and the Netherlands thought that they were attending an online meeting with member of the Russian opposition Leonid Volkov. Later on, however, it turned out that the person attending the meeting was not Volkov, but that someone was imitating him by using **deepfake technology with the aim to discredit** Volkov[64]. Accounts of the incident have indicated the involvement of a foreign intelligence agency behind it.

The examples above supposedly originate from the hands of different sectors – one appears to have been created by disgruntled political supporters, while the other one appears to have been created with a deeper political agenda. In both cases, however, deep-fakes offer new pathways for engaging with and manipulating public opinion, displaying enormous potential to undermine trust in public representatives and institutions, and to cause reputational damage. The examples also show the complexity of the landscape, in which deep-fakes operate and the parties involved – public and private actors; social media platforms and their users; technology providers. In the following section we reflect on the AIA provisions outlined above, and on how they interplay with the examples.

5.2.3. Analysis

Two of the identified divergences between public and private obligations in relation to deep-fakes relate to law enforcement – 1) Art. 52(3)2 AIA frees them from the obligation to disclose that manipulated content is at play, in contrast to the labelling requirement for users from the private sector; and 2) while Annex III 6(c) AIA triggers stricter rules on using software to detect deep-fakes in investigations, Art. 52(3) AIA treats the very systems whose detection is sought after as low risk (linked to transparency obligations for private sector actors). Both cases show incoherence in the underlying levels of risk the provisions seek to address and in their assessment, as well as lack of clarity regarding the type of information the transparency obligation entails.

The examples in the previous section are both facilitated by digital technologies (online platforms) that allow deep-fakes to travel across jurisdictional boundaries and have far-reaching consequences for their targets on political and personal level. In both instances, the AI users appear to deploy the deep-fakes with malicious intent, circumventing (potential) labelling or disclosure of the manipulated content, and affecting fundamental values and democratic structures. Yet, the

---

59 Galson, W., 'Is seeing still believing? The deepfake challenge to truth in politics', The Brookings Institution., January 2020.

60 Overdub: Ultra realistic text to speech voice cloning, Descript.

61 The Washington Post, Another fake video of Nancy Pelosi goes viral on Facebook, 3 August 2020

62 EUobserver, 'Deepfakes' - a political problem already hitting the EU, 2021.

63 The Guardian, European MPs targeted by deepfake video calls imitating Russian opposition, 22 April 2021.

64 NOS. Veel onduidelijk over 'deepfake-gesprek' van Kamerleden met medewerker Navalny, 24 April 2021.

proposal does not appear to consistently follow through on the intent behind the deployment of deep-fakes, nor on the potential harms. It arbitrarily bundles these with transparency obligations for private sector users, removes them for law enforcement, and then strengthens the requirements again for the deployment of software for their detection by law enforcement. The rationale behind the risk approach is difficult to follow. Further, the burden of disclosure falls on the user and not on the provider of the system. We see the same with regard to the detection of deep fakes – it is the use of law enforcement that is considered high-risk, not the creation of the system as such. Technology appears to be considered here as being neutral.

These, however, are not the only inconsistencies in the division of obligations between public and private actors with regard to manipulative AI systems. The divergence between Art. 5(1)(b) AIA and Art. 5(3) UCPD – the address of unintentional (commercial) practices in the latter and lack of it in the former – continue the red thread. Here, however, intent is not only considered, it is paramount for the applicability of the AIA ban. Taken together with the previous paragraph, in which harm or intent do not play a role, the AIA paints an incomplete picture in how and why manipulative AI practices should be addressed, by whom and what considerations have been taken into account to shift the burden of detection towards law enforcement.

5.3. Social scoring

5.3.1. AIA divergences

	Public sector	Private sector
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 gives explicit prohibition of sources of certain data 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

Art. 5(1)(c) AIA prohibits:

> *the placing on the market, putting into service or use of AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to either or both of the following:*
>
> *(i) detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected;*
>
> *(ii) detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;*

The provision targets the use or sale of AI systems by 1) public authorities ('by or on behalf of'); that 2) produce scores of 'trustworthiness' of individuals; 3) based on social behaviour or *personal* characteristics; 4) which lead to either disproportionate or unjustified treatment of (groups of) individuals, or to treatment that is proportionate and justifiable but occurs in an unrelated 'context' to the one from which the input data originates.

Art. 5(1)(c) AIA as such is an **explicit ban for the use of social scoring in the public sector**. A potential private involvement is sanctioned only if it happens ***on behalf of*** **public authorities, leaving private sector actors outside of the scope of the provision**. Not inserting social scoring practices by private sector actors in the prohibition of Art. 5(1)(c) AIA means that those fall under the high-risk regime of Title III AIA. Accordingly, private sector scoring systems that assign persons to educational institutions or such that determine the enjoyment of essential private services (see 3. (a) and (b), as well as 4 (b) and 5 (a), (b) and (c) of Annex III AIA), to name just a few, are subject to the *ex-ante* and *ex-post* requirements of Chapter 2 and Chapter 3 of the AIA in order to be placed on the EU market or used.

As it is becoming increasingly difficult to draw a clear line between public- and private-facilitated use of AI, these diverging obligations for social scoring are problematic in a number of ways. To begin with, neither Art. 5(1)(c) AIA nor other parts of the AIA offer more clarity on what uses of AI systems are considered **to take place *on behalf of* public authorities** or how that relationship is to be further structured and elucidated. The importance of the requirement, however, is not to be underestimated, since digital technologies deployed by public actors are **mostly developed by the private sector**. It is unclear **how and to what extent off-the shelf solutions** – not explicitly developed for administrators but used by them – relate to Art. 5(1)(c). This is further problematic regarding the many social media applications (Twitter, LinkedIn or Facebook) which are privately own and run, but often play a palpable role in **background checks for public employment**. Such applications are not strictly regarded as social scoring though clearly used to make a pronouncement on the person's aptness or trustworthiness. In addition, the layer of data brokers, advertising intermediaries and recommender systems intertwined with big tech platforms and relying on their own scoring processes to nudge online behaviour, have proven to be disruptive of democratic processes and detrimental to economic markets[65]. Yet, they are also out of the provision's scope.

Another issue in the assignment of diverging obligations between public and private actors in the context of social scoring is related to data. For the ban of public sector social scoring Art. 5(1)(c) AIA states that personal and/or personality characteristics and social behaviour are prohibited data sources. For the high-risk social scoring of private parties, however, Art. 10 AIA envisions only data quality and data governance requirements. **From a data perspective, the grounds for prohibition for the public sector are more detailed than those on use for private actors**. The latter omission can be partially explained by Recital 44 AIA, which explicitly refers high-risk AI to the requirements of other EU legal regimes, including data protection. Recital 44 AIA ensures as such that the provisions of the Act are applied only in accordance with other existing law, and clarifies that the AIA cannot be understood as an additional ground for the processing of personal data. It is questionable, however, whether this 'generic statement of compatibility'[66] between the AIA and the GDPR solves all data use issues brought about by private sector social scoring. It is unclear, for instance, whether an application of Art. 5(1)(c) would be triggered in the case **of scoring developed with private sector datasets of augmented administrative data**[67].

In the following, we portray examples of social scoring in the public and private sector, before discussing the regulatory divergences thereon in the final sub-section.

---

65 Hildebrandt, M. The Proposal for an EU AI Act of 21 April 2021. Brief Commentary. 19 July 2021.

66 Codagnone, C. et Al., Identification and assessment of existing and draft EU legislation in the digital field, Study for the special committee on Artificial Intelligence in a Digital Age (AIDA) , Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2022, p. 62.

67 Veale & Borgesius, 2021, p. 102; Ebers et al., 2021, p. 592.

5.3.2. Examples of use of social scoring from the practice

In the Netherlands, the Dutch government developed legislation on *System Risk Indication* (SyRI), aimed at detecting fraud, including social benefits fraud, allowance fraud and tax fraud, by looking into profiles that are 'abnormal' or 'unlikely'[68]. SyRI **took into account zip codes and was deployed in 'problem' neighbourhoods** in Rotterdam, Haarlem and Capelle aan den IJssel. The Dutch court ruled that the legislation regulating SyRI's use is in violation with higher law, as it did not comply with Art. 8 of the European Convention on Human Rights (ECHR), protecting the right to respect for private and family life, home and correspondence. The court also found that **the file linkage used in SyRI meets the definition of profiling** within the meaning of Art. 4 (4) of the GDPR. The court further established that **SyRI's data use was not proportional**. Data sources that qualified for processing in SyRI were data about work, housing, administrative measures and sanctions, taxes, movable and immovable property, civic integration, education, pensions, debt burden, social benefit, permits and exemptions, health care insurance, trade and identifying data including name, address, city, postal address, data of birth, gender and administrative characteristics[69]. Moreover, SyRI was implemented **without a transparent process** that shows citizens what happens with their data[70]. The Dutch government refused to share with the court information on the risk indicators used, and on how the risk model worked, insisting it to be 'secret'. As a result, though the court did not get sufficient insight into what SyRI entails and how the algorithms used in the model work, it stressed the importance of transparency, as using a risk model in this context brings the risk of discriminatory effects by, for instance, incorrectly attributing certain characteristics to people[71].

Meanwhile, several MS governments (have) run experiments aimed at detecting fraud or problematic behaviour that could be classified as 'citizen scoring'[72]. In Denmark, for instance, a profiling system was developed with the aim to detect children in vulnerable families. The systems was based on a points system and **included indicators such as the mental health of parents, unemployment, missed medical appointments and divorce**. However, after critical media coverage and the decision of the Data Protection Authority to deny permission for the continuation of the project, the pilot was stopped[73]. In Spain, an algorithm is used to **determine whether people are eligible for financial aid** to cover their electricity bills. *Civio*, an investigative newsroom and citizen lobby, discovered that the software did not work properly and that it systematically denied financial aid to eligible people. In addition, the Spanish government refused to share the source code of the algorithm, basing itself on copyright considerations[74]. Further, the French Ministry of Finance uses **machine learning to detect tax fraud**. The ministry also introduced an amendment to the budget law that allows it to scrape data from social networks, auction platforms and advertising websites to detect tax fraud. The Data Protection Authority has criticized this plan[75].

Insurance companies also develop tools to track the behaviour of customers, and give a discount on their insurance policy in return. For example, the Royal Dutch Touring Club ANWB offers a car **insurance discount** to customers who are willing to download an **app that tracks their driving style**[76]. In a similar fashion, the Dutch insurance company ASR offers a discount to customers if they

---

68 Algorithm watch. Automating Society Report 2020.

69 ECLI:NL:RBDHA:2020:1878, SyRI court ruling. Rechtbank Den Haag. 6 March 2020.

70 Algorithm watch. Automating Society Report 2020.

71 ECLI:NL:RBDHA:2020:1878, SyRI court ruling. Rechtbank Den Haag. 6 March 2020.

72 Algorithm Watch. Personal Scoring in the EU: Not quite Black Mirror yet, at least if you're rich, 2019.

73 Algorithm Watch., Denmark - Automating Society Report 2020.

74 Algorithm Watch. Spain - Automating Society Report 2020.

75 Algorithm Watch. France - Automating Society Report 2020.

76 ANWB Veilig Rijden Autoverzekering, ANWB website (in Dutch).

download an app that tracks their activity. If customers lead an active lifestyle, they can **earn more points** and in return get a discount on their health insurance[77].

We found further scoring examples in the finance sector. Usually, a pre-requisite for obtaining a loan from financial institutions is a credit score that attests to the creditworthiness of a consumer. However, some people are not able to obtain a loan from traditional financial institutions due to a short credit history or the lack of an official form of identification[78]. As an alternative, a growing number of companies is offering the option to **assess someone's creditworthiness by using AI and alternative data** such as mobile data and social media data[79]. For example, *Credolab*, a company based in Singapore and active in 15 countries, offers an **alternative credit scoring application based on smartphone and web data**[80]. And *Tala*, a US fintech active in Kenya, the Philippines, Mexico and India offers a financial service to the unbanked – people without an account at an official financial institution[81]. By using ML they analyse **android device data and behavioural data** to determine a customer's eligibility[82]. How a decision is made based on the alternative data and ML models, and which steps are necessary to build up a good credit score, is unclear[83].

5.3.3. Analysis

The examples above **question the robustness of the regulatory choice behind prohibiting social scoring for public actors vs. designating it high-risk for private sector applications**. Similar to the ban on manipulative AI and the provisions related to it, we notice an inconsistency in the assignment of risk levels and their assessment. While we acknowledge that the provisions of the high-risk regime are a first step in the right direction when it comes to setting-up or monitoring a high-risk AI system, they **do not do justice to the harm inflicted by private social scoring practices to individuals, and do not provide redress mechanisms that can tackle the power imbalance emerging thereby**. The transparency and information provision to users in Art. 13 AIA does little to remedy that. It is doubtful individuals would be better off if those deploying the system are able to comprehend the '[…] characteristics, capabilities and limitations […]'[84] of the scoring system when the latter **continuously accompanies and evaluates citizens' conduct**, and thereby scores their health, commitment, reliability, etc. and/or any imaginable skill in order to provide an essential service such as insurance or identification. It is unclear why such a state of affairs is less threatening to individual and collective rights and freedoms than social scoring conducted by public authorities.

Symptomatic for the misalignment of risk between the ban of public social scoring vs. allowing it under the high-risk regime for the private sector is also the missed opportunity to link it to the risk levels of Art. 35 GDPR. It remains unclear whether the fact that the AI system is qualified as high-risk triggers the high-risk processing requirements of Art. 35 GDPR as well. It is also challenging to establish how consent rules would apply. The draft ePrivacy Regulation envisions among others updated consent rules for the processing of electronic communications data (see Art. 6 and Art. 7 draft ePrivacy Regulation) and an exhaustive list of purpose exceptions therefrom. However, the purpose of training (Art. 10 AIA) is currently not part of the consent exception list in

---

77 Kom in beweging met a.s.r. Vitality, ASR website (in Dutch).

78 Agarwal, S., Alok, S., Ghosh, P. & Gupta, S., Financial Inclusion and Alternate Credit Scoring: Role of Big Data and Machine Learning in Fintech, *Indian School of Business*. SSRN, 2021.

79 OECD. 'Personal Data Use in Financial Services and the Role of Financial Education', 2020.

80 Digital Credit Scoring with Alternative Data, Credolab website.

81 The unbanked - Global Findex, World Bank website.

82 Data Ethics, Tala website.

83 OECD. 'Personal Data Use in Financial Services and the Role of Financial Education', 2020.

84 See Art. 13(3)(b) AIA.

the draft ePrivacy Regulation. The rules supposed to govern social scoring by private actors are as such not only misaligned with the risk levels of the ones underlying the social scoring ban for public authorities; they show also risk-based inconsistencies with data protection provisions.

The latter leads us back to the second divergence in this context – the one related to **data, datasets, and their use by public or private actors in the context of social scoring**. The examples above include the use of health data, proxies of personal data, behavioural and meta data. In that regard it is worrying that the AIA does not include a more detailed reference to the obligations and prohibitions of the GDPR, and how those can effectively retrain social scoring applications. A case in point is the relationship **between the AIA's provisions** on **private sector social scoring and Art. 22 GDPR** on profiling of natural persons. Art. 22 GDPR prohibits both the automated processing of data to evaluate certain individual traits or characteristics, as well as automated decision-making without human involvement. While those requirements would be triggered in most of the private sector social scoring examples outlined above, Art. 22 GDPR and (Recital 71 GDPR) apply neither in cases of semi-automated or hybrid AI systems that involve both human and automated decision-making processes, nor when the person affected by the scoring or profiling is not the data subject.

5.4. Biometrics

5.4.1. AIA divergences

	Public sector	Private sector
Biometrics	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 wrt their use of ERS and BCS	Art. 52(2) AIA envisions transparency obligations for the use of ERS and BCS by private sector actors

AI-enabled biometric systems were one of the primary regulatory targets of the AIA. In fact, the Act explicitly refers to three types of biometric systems – emotion recognition systems (ERS), biometric categorisation systems (BCS) and remote biometric identification systems (BIS). In the case of the latter, the proposal further distinguishes between 'real-time' and 'post' BIS. While real-time BIS accomplish their task instantaneously or without a significant delay based on live (or near-live) material, the 'post' BIS are run after a significant period of time based on data from private devices and/or CCTV cameras. The AIA thus assigns them different societal gravity and therefore regulatory approaches. As will be explained below, **some biometric systems are prohibited for public actors, but categorized as high-risk in the use of the private sector**. Yet other AI systems underlie (mere) transparency obligations for private actors, while the same transparency obligations are lifted for law enforcement.

Prohibited practices and high-risk AI systems

Art. 5(1)(d) AIA prohibits

> *the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives […].*

By dedicating the prohibition to law enforcement, the **AIA makes it lex specialis to the LED**. Art 5(1)(d) grants, however, **three exceptions to the use of real-time BIS in the case of public security concerns** relating to: (i) time-sensitive search for victims of crime, (ii) prevention of terrorist

attacks, or (iii) the detection of suspects of a criminal offence concerned by a custodial sentence or a detention order for a period of at least three years.

The AIA understands real-time BIS as such systems that capture, compare, and identify individuals 'instantaneously, near-instantaneously or in any event without a significant delay'. Art. 3(36) and (37) AIA add to Art. 5(1)(d) AIA by clarifying that it is a person's biometric data as such that is being captured and compared with biometric data contained in a database. Biometric data in turn is personal data obtained through 'specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data'. Recital (7) AIA further clarifies that 'biometric data' is to be understood in line with Art. 4(14) GDPR and 3(13) LED. In AI systems that perform real-time identification, the latter occurs without prior knowledge of the user of the AI system whether the sought after person will be present.

When used by law enforcement, real-time BIS are deemed particularly intrusive on individual rights and freedoms, given the immediacy of the identification and the limited to none available redress mechanisms individuals can make use of in that instance for further checks or corrections. In particular, real-time BIS require a great amount of sensitive data to be processed in order to perform an identification, severely interfering with data protection and privacy rights. They further endow the feeling of being subjected to constant surveillance and may dissuade the exercise of the freedom of assembly and other fundamental rights. The prohibition thus intends to prevent the disproportionate use of real-time BIS in the prosecution of minor offences or in cases of public protests.

In contrast, the **use of real-time BIS by the private sector is not banned but considered high-risk** and handled by Title III of the AIA, of which Art. 6(2) and its corresponding Annex III are here particularly relevant. BIS that qualify as high-risk must comply with the *ex-ante* and *ex-post* requirements of Chapter 2 and Chapter 3 of the AIA in order to be placed on the Union market or used. The pre-market check requires providers to set up a risk management system (Art. 9 AIA), to follow strict rules on data training, validation and testing (Art. 10 AIA), to ensure transparency and provide necessary and adequate information to users (Art. 13 AIA), to facilitate appropriate human oversight in order to minimise risks to health, safety and fundamental rights (Art. 14 AIA), and to ensure that the system's design is accurate, robust and following cybersecurity considerations (Art. 15 AIA). Further, BIS have to fulfil *ex-ante* conformity procedures (Art. 16-19 AIA) and in line with Art. 43(1) AIA undergo conformity assessment by an independent body unless common specifications or harmonized standards exist.

Once a BIS has complied with the ex-ante requirements, it can be placed on the market in line with other Union legislation. As an ex-post check Art. 61 AIA envisions market surveillance and supervision of such systems by national authorities designated by Member States. However, all *ex-ante* and *ex-post* requirements are broadly formulated and leave the details around the actual standards to the European standardisation bodies. The latter means that the **success of the provisions outlined above will depend to a very large extent on harmonised standards from the private sector**. For instance, the formulated data requirements in Art. 10 AIA aim to prevent AI systems from becoming the source of discrimination prohibited under Union law, but do not stipulate the prohibited forms of bias or the bias mitigation measures that should be applicable. The same is true for the information provision and transparency obligations in Art. 13 AIA. **The appropriate degree of transparency for private sector users of BIS is left open**.

Low risk and biometric categorisation systems (BCS)

Biometric categorisation systems – those systems that biometrically cluster individuals according to categories like 'sex, age, hair colour, eye colour, tattoos, ethnic origin or sexual or political orientation, on the basis of their biometric data' –, or emotion recognition systems (ERS) that

according to Art. 3(34) AIA are intended to assess or infer the emotional state of individuals on the basis of their biometrical data, are neither prohibited nor featured in the list of high-risk AI systems. Consequently, **BCS and ECS fall in the category of AI systems that pose limited risk** and are therefore subject to the limited **transparency and information obligations** of Art. 52(2) AIA for both public and private actors, **with the exception of law enforcement**.

Before discussing the implications of the AIA's diverging approach towards the different biometric systems, we present below examples of their use by public and private actors.

5.4.2. Examples of the use of biometric AI systems from the practice

BIS systems

In the US dozens of local governments and cities are issuing bans for the use of facial recognition in the public sector[85]. Though the bans' application vary from jurisdiction to jurisdiction, a number of them apply beyond law enforcement practices. Proposal legislation for similar bans at federal level[86] is currently being considered in both houses of the US Congress. In stark contrast, however, is the technology's increased use at US airports to verify the identity of passengers. Checking incoming passengers by means of facial recognition technology is currently carried out at 17 US airports, while 32 have the capacity to do so for outgoing travellers. Cooperating with additional airports, US Customs and Border Protection intend to roll out the program together with the Transportation Security Administration, and travel security companies[87]. Although the AIA is scoped to target deployment of AI systems in Europe, such examples act as signposts for regulatory divergences around the use of BIS in Europe.

A European example of use of facial recognition technology is a pilot for identity verification conducted in July 2021 in a public-private partnership between the Dutch government and Schiphol airport. Travellers who participated in the pilot were able to pass contactlessly through checkpoints at Schiphol without needing to show their passport and boarding pass, because they were being recognised by their face[88]. In 2017 and 2018 Paris-Charles de Gaulle airport modernised its Automated Fast Track Crossing at External borders system by introducing face recognition technology[89]. There are concerns about the reliability of the use of face recognition technologies for such purposes. For example, the use of a facial recognition system at Brussels Airport was stopped because there were persistent errors and it led to inefficiency[90]. In addition, research has shown that facial recognition applications can be less accurate for certain people like women, and people with darker skin tones[91].

A similarly controversial example yet from the private sector and with a different magnitude is Clearview AI – the facial recognition company that scraped about 10 billion facial images from the Internet without the consent of individuals, They used the images to develop an identity-matching service (providing names and other available information), which was eventually sold both to law

---

85 Wired, Face Recognition Is Being Banned—but It's Still Everywhere, 22 December 2021.

86 The Washington Post, Civil rights groups push Biden administration to take stand against facial recognition, 17 February 2021.

87 Wired, Face Recognition Is Being Banned—but It's Still Everywhere, 22 December 2021.

88 Facial recognition pilot for departing travellers, Schiphol website and Travel with facial recognition, Schiphol website.

89 Smart gates for Paris Orly & Charles de Gaulle, Thales website.

90 Dumbrava, C., 'Artificial Intelligence at EU borders - overview of applications and key issues', EPRS, European Parliament, July 2021.

91 Dumbrava, C., 'Artificial Intelligence at EU borders - overview of applications and key issues', EPRS, European Parliament, July 2021.

enforcement agencies and to private actors for market surveillance purposes[92]. What the company branded as a crime-solving tool contains almost any image posted online, providing both law enforcement and corporate actors such as banks and shopping centres with a massive database exceeding any traditional government dataset[93]. The data protection authority of France recently stated that Clearview AI has breached the GDPR[94]. The UK data protection authorities[95], as well as the Canadian Privacy Commissioner[96] have also addressed Clearview's practices and found the app to be inappropriate and posing risks to individuals. The authorities have further commented on the way in which Clearview AI obtained its data, and found the accumulation of images to be indiscriminate, unreasonable and inappropriate, especially since the original purpose of the individual images was unrelated to law enforcement or market surveillance purposes.

Another example concerns a Dutch supermarket that deployed facial recognition technology on top of store cameras to scan the faces of shop visitors, and compare them to its database of shop-lifters and other individuals banned from entering the premises[97]. The Dutch Data Protection Authority has issued a formal warning to the owner, stressing that entering a store does not equal giving individual consent, even when a warning sign has been positioned at the door.

ERS and BCS

Emotion recognition technology is increasingly used for a number of purposes in private sector applications. A case in point is *Realeyes*, a UK start-up with an office in Budapest, that developed a technology using computer vision and machine learning to measure audience attentiveness and engagement when watching a video or an ad[98]. *Behavioral Signals* in turn is a provider of technology that uses emotion recognition AI and voice data to match customer service agents or sales agents to customers. According to the website of *Behavioral Signals*, a bank and an energy provider in the EU use its technology to optimise their respective debt collection processes[99]. *WeSee*, a British-Slovenian company offers a technology that uses emotion recognition to detect insurance fraud. Using deep learning to assess the reactions and responses of an individual to a set of questions in real-time, the application immediately delivers an assessment to the insurer. The company's website states that the application can detect emotions in real-time by monitoring micro-expressions, eye movement, gaze, speech patterns and voice patterns[100]. The use of these applications is regarded as problematic because it runs the **risk of misinterpretation, inaccuracy or bias**. A study by Fölster, Hess and Werheid, for example, found that facial age affects the decoding of emotional expression,

---

92 Gonzalez Fuster, G. and Nadolna Peeters, M., Person Identification, Human Rights and Ethical Principles: Rethinking biometrics in the era of Artificial Intelligence, EPRS, European Parliament, December 2021.

And Techcrunch, France latest to slap Clearview AI with order to delete data, 16 December 2021 And TechCrunch, France latest to slap Clearview AI with order to delete data, 16 December 2021.

93 Rowe, E. A., Regulating Facial Recognition Technology in the Private Sector. *Stanford Technology Law Review*, 24(1), 2020.

94 Techcrunch, France latest to slap Clearview AI with order to delete data, 16 December 2021 and Reconnaissance faciale: la CNIL met en demeure CLEARVIEW AI de cesser la réutilisation de photographies accessibles sur internet, CNIL, 16 December 2021.

95 TechCrunch, Clearview AI told to stop processing UK data as ICO warns of possible fine, 29 November 2021.

96 Announcement: Clearview AI ordered to comply with recommendations to stop collecting, sharing images, Office of the Privacy Commissioner of Canada.

97 Dutch DPA issues Formal Warning to a Supermarket for its use of Facial Recognition Technology. European Data Protection Board website.

98 About us - Realeyes, Realeyes website and Realeyes raises $12.4 million to help brands detect emotion using AI on facial expressions, VentureBeat website.

99 AI-Mediated Conversations, Behavioral Signals website.

100 Emotion Recognition Could Eliminate Insurance Fraud, WeSee website.

because there is a difference in the way older and younger people express emotions[101]. In addition, AI-based emotion detection technology can have a serious impact on minorities and vulnerable people, including people that have anxiety or certain tics[102].

Another controversial example that was recently discontinued is *HireVue's* application that uses emotional recognition and facial recognition technology to screen potential job applicants. After public outcry over the use of this application and a complaint to the FTC by US-based Electronic Privacy Information Center that HireVue failed to guarantee fairness and of using algorithms that cannot be audited, HireVue decided to discontinue with a part of its software: using emotion recognition in a video to identify certain characteristics[103]. However, the company still uses language of applicants to assess their employability applicant[104].

In addition, automotive companies like BMW, Kia and Porsche increasingly use emotion recognition technology to monitor whether a driver is alert and has her eyes on the road while driving[105]. In January 2020, the European Data Protection Board published guidelines on the processing of personal data in connected vehicles. The guidelines state that biometric data may be used, among other things, for the following purposes: to access a car, to authenticate the driver and/or to enable access to the profile of a driver. The guidelines, however, do state that personally identifying information cannot leave the car without the consent of the user[106].

5.4.3. Analysis

The above-mentioned examples of real-time BIS in airports illustrate that these **AI systems are often developed in public-private partnerships**. Therewith, the boundaries between public and private actors and their respective domains of operation become more blurred, creating difficulties in establishing at which point the involvement of a public party begins and ends, and where exactly a private party steps is and with what ramifications. Further, reviewing the examples presented above does not render it defendable to 1) hold law enforcement to a higher standard than other public or private actors in the application of real-time BIS; and 2) to differently measure similar risks for fundamental rights and values in the deployment of such systems.

The rationale behind all those regulatory choices appears to be supported by the purpose and immediacy (e.g. the time or *real-time*) of the identification as the main factors of intrusiveness that need to be countered. However, **it is the *use* and its implications, and not the AI application's timeline that should be central in the prohibition**, unless the weight of the time component is empirically proven. Note that issues of scope (as in the case of Clearview AI with identifying millions of individuals), consent (the data obtained without individual knowledge as among other in the Dutch grocery shop example), the technical architecture needed to successfully run the system (centralised biometric databases), combined with capturing and processing the data of every single person in a respective space, **do not differ in real-time BIS employed by public or private actors. In either case, the AI system affects individuals who have no reason to be monitored**, with the palpable difference, however, that private parties are not bound by considerations of common

---

101 Fölster, M., Hess, U., & Werheid, K. 'Facial age affects emotional expression decoding'. Front. Physchol, 2014.

102 Dumbrava, C., 'Artificial Intelligence at EU borders - overview of applications and key issues', EPRS, European Parliament, July 2021.

103 Wired, Job Screening Service Halts Facial Analysis of Applicants, 12 January 2021.

104 Financial Times, Emotion recognition: can AI detect human feelings from a face?

105 Vice, Car Companies Want to Monitor Your Every Move With Emotion-Detecting AI, 27 July 2020 and The Sun, BMW's new X5 SUV has a camera that alerts you when your eyes wander off the road, 2 October 2018.

106 European Data Protection Board, Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications, 28 January 2020.

good, public values or the checks and balances present in democratic societies that public authorities attend to.

The **AIA's approach to the ban in Art. 5(1)(d) also differs from that in the GDPR**, which does not distinguish between public and private data controllers. By creating a general category, the GDPR recognises that data technologies create new powers for industry actors to exert control at population scale and expose fundamental rights to new forms of intrusion. By omitting to apply the same rationale to AI-facilitated BIS used by private actors and other (than law enforcement) public actors, the AIA shows itself as regulatorily incoherent.

In addition, the prohibition of **Art. 5(1)(d) AIA focuses on BIS in 'publicly accessible spaces'**. Those are according to Art. 3 (39) AIA 'any physical place[s] accessible to the public, regardless of whether certain conditions for access may apply (emphasis added)'. A number of the applications mentioned above, however, such as Clearview AI (the latter of which was eventually used by law enforcement across the globe) are implemented **online**. Excluding online spaces from the scope of the prohibition seems to be in direct contradiction to Recital (6) AIA, which when clarifying the notion of AI systems explicitly refers to their effects '[…] in a physical or digital dimension.' Without a particular justification of why online spaces should not be seen as publicly accessible spaces, the AIA takes insufficiently account of specific harms that stem from the online deployment of AI biometric systems. We consider this incoherence to be both **weakening the prohibition of Art. 5(1)(d) AIA for the public sector, as well as benefitting private sector actors who already have the first-mover advantage online**.

Further, the rationale of embedding BCS and ERS in the low-risk part of the AIA and submitting them to mere transparency obligations is questionable. There is no significant difference between the threats to the freedoms of speech and assembly (and other fundamental rights foundational to a democratic order) posed by BIS on the one hand, and such effects produced by BCS/ERS in (semi-) public spaces. As seen in the examples presented above, some of these systems are already in use in vital services such as banking and insurance, measuring and profiling individuals. Further, the results of both BCS and ERS are often at the root cause for biased decision-making, processing and clustering of individuals according to characteristics that are grounds for discrimination prohibited under Art. 21 EU Charter of Fundamental Rights. **Against this background, the lack of transparency obligations for law enforcement in Art. 52(2) AIA - the only attempt at individual protection that the AIA actually provides – as opposed to its applicability to private sector actors that deploy BCS and ERS is worrisome**. Systems with such intrusive capacities should be subject to checks and balances, especially when in the hands of public actors. Here again, the AIA's incoherent risk assessment approach makes itself noticeable.

6. Key findings and discussion

In this chapter we summarise and where possible generalise the findings from the three analyses sections above on how the AI Act diverges in its obligations towards public and private actors in the three areas of application (manipulative AI systems, social scoring and biometric AI systems).

6.1. AIA divergences - treating similar AI systems differently depending on the user

The identified divergences relate to **treating similar practices** (uses of AI systems) **differently depending on the actors that deploy them, and to placing similar systems in different risk categories**. We see the former in the dichotomy of public vs. private sector obligations in relation to social scoring, as well as in relation to the prohibition of real-time BIS for law enforcement. The latter in turn is especially prominent in designating systems that law enforcement uses for the detection of deep fakes as high-risk (Annex III, 6(c) AIA) vs. designating deep fakes low risk in general.

Historically, the difference in treatment is somewhat understandable – it is surveillance by the state individuals need protection from, and the examples from the public sector in fraud-detection, social benefits allocation or facial recognition misuse have fed the regulator's appetite. However, this neat division of responsibilities is not only impractical in the context of AI systems and AI-facilitated services; also in terms of data and lifecycle management of that data there is a palpable entanglement of IT industry data and AI tools, and platforms and public sector services[107]. The separation of private and public actors' AI practices also seems less and less defendable, as the **risk levels associated with AI use by either public or private actors do not differ in the power asymmetry they create towards the individual**.

The widely used cross-context social scoring applications by the private sector oversee financial flows, credit card records, critical infrastructure such as transport, welfare support or insurance. In the examples presented earlier, these applications score and profile individuals' daily activities and lifestyle, decide how and whether they get access to vital services, and at what pricing. They operate thereby in the 'public sphere' – a space in which not only commercial but also political and social action takes place; a space that harbours matters of public concern[108]. As such, **social scoring AI applications by the private actors affect the population at a general, foundational level**. The latter is important, as it goes beyond legal or administrative solutions to private sector faux pas – it **elevates (AI) service providers to public service providers**. The implication thereof is that while companies have profit goals in mind that prevail over public goals or responsibilities, their presence and modus operandi in the public sphere influence political agendas and policy[109], as well as public investments and academic research[110]. It also illustrates the difficulty of drawing a clear line between public values and activities and private ones, or in remedying an imbalance between those values

---

[107] See Delipetrev, B., Tsinaraki, C., Nepelski, D., Gomez Gutierrez, E., Martinez Plumed, F., Misuraca, G., De Prato, G., Fullerton, K., & Craglia, M., and Duch-Bro, N., 2020. 'AI Watch 2019 Activity Report,' JRC Working Papers JRC121011, Joint Research Centre (Seville site).

[108] Taylor, L. Public actors without public values: legitimacy, domination and the regulation of the technology sector. *Philosophy & technology*, 34(4), 2021, pp.897-922.

[109] The lobby spending of big tech in Brussels far outstrips that of any other sector, see European Association for Digital Transition, Big tech lobbying in Brussels: A lot of money. Too much influence?, 4 March 2021.

[110] See for instance the @fundingmatters community, that tries to show the influence of big tech on academic research. Statements about the corporate funding of academia, Funding Matters and The Dark Side of Big Tech's Funding for AI Research, Wired.

and activities by means of a regulatory approach based on their functional separation[111]. As values and activities converge, it becomes even more difficult to define responsibilities and match the checks and balances we have installed in the offline worlds. Such an approach **does not appear to be grounded in context-related conceptions of AI-induced harms or risks**.

In a similar fashion, the discussed examples of **real-time BIS designed and deployed by private actors are problematic not only from a privacy or data protection perspective**, which are always implicated[112]. Next to biased or discriminatory outcomes of such systems, the less obvious repercussions are the erosion of a right to anonymity; the far-reaching consequences for children's rights (whose unassuming parents share pictures of on social media platforms, which are then scraped for training computer-vision AI-systems by companies); the cumulatively chilling effects for the freedoms of speech, expression (including political orientation[113]), assembly, and religion, among others. The silent omnipresence of such systems in public and semi-public spaces, as well as on the Internet points to a real and overarching monitoring of individuals that **however does not undergo any legal justification or checks of purpose, necessity and proportionality**, since it is not performed by state bodies.

The divergences in relation to placing **similar types of systems in different risk categories** are further significant as they undermine the AIA's purpose in harmonising AI-related provisions and creating more legal certainty. Neither the wording of the provisions, nor the AIA's Recital provide conclusive information on how these risk levels have been assessed and on the basis of what criteria and thresholds. To a large extent, these considerations also apply to the distinction between real-time and post BIS, as well as to BIS systems in general and ERS/ BCS. The risks and harms associated with their deployment are not clearly delineated; certainly not clear enough to render some of them high-risk and other low-risk. For instance, the AIA's explanatory memorandum speaks of calculating risks by 'taking into account the impact on rights and safety'[114], while the Recital refers to considerations of '[…] intensity and scope of the risks […]'[115] generated by AI systems. Yet these categories remain abstract and general and contribute to the Act's regulatory incoherence.

6.2. Regulatory (in)coherence with EU law

Besides the number of public vs. private regulatory divergences found in the AIA, a follow-up step is to look at whether these divergences lead to (in)coherence with already existing or proposed EU law, and how. As stated in the methodology section, this step is of relevance because the AIA will not land in a regulatory vacuum: rather, it will interact with generic or sector-specific legal frameworks. Accordingly, though it is beyond the scope of this study to analyse the entire legal landscape of EU regulation, it is fruitful to go one step beyond the identified divergences and connect them to the rationales of obligations and risk-levels harboured within related Regulations and Directives. Throughout the report, we have already made such references where appropriate.

Reflecting on the potential lack of regulatory coherence between the AIA and other sources of Union law we divide these into **issues of scope and procedure** of the proposed Act. By scope we mean the substantial elements in the Act that might lead to diverging interpretations. Think for instance of the definitions of an AI-system 'user' in the AI Act and a 'data controller' in the GDPR. By procedure,

---

111 Taylor, L., Public actors without public values: legitimacy, domination and the regulation of the technology sector. *Philosophy & technology*, 34(4), 2021, pp. 897-922.

112 Smuha, N. A., et al., 2021, p. 24.

113 Kosinski, M., Facial recognition technology can expose political orientation from naturalistic facial images. *Scientific reports*, 11(1), 2021, 1-7.

114 Explanatory Memorandum to the AIA (8).

115 Recital (14) AIA.

in turn, we mean the procedures and obligations proposed in the AIA and how they might diverge from procedures and obligations laid down in other EU law. In this section we aim to list them more structurally, with the general line or argumentation being that there is a divergence identified in the AIA between public and private actors, which leads to a policy option to either solve this divergence within the AIA or to relate this divergence to other legal frameworks. This in turn can lead to a lack of regulatory coherence, which we will briefly point out.

6.2.1. Scope

Those related to scope are threefold. The first one relates to the narrow scope of the harm requirement in Art. 5(1)(a) and (b) ('physical or psychological harm') and its relation to general Union law, which as a rule of thumb provides broad harm definitions and leaves those to be filled by the MS. By explicitly defining the harm, the AIA intends to complement the UCPD and fill gaps left by the latter. However, it is **questionable how and whether such an approach is compatible with the AIA's horizontal and general nature**, which should be applicable across sectors and contexts. Further, as known from tort liability law, **to demonstrate or prove individual harm** even when the latter is probable is **nearly impossible**, though the proof will be required by Art. 71(3)(a) AIA to impose a fine[116].

The second issue of scope also refers to divergences between the prohibition of manipulative AI systems and the rules of the UCPD. By means of the strict requirement of intent, and by excluding persons with a vulnerability different than such based on age, physical or mental disability, Art. 5(1)(a) and (b) AIA create loopholes for the private sector. Making the prohibition **exclusively dependant on the intentionality** with which the AI system is deployed, **excludes cumulative harms**, which build up over time through the interaction of the system with its contextual setting. However, without an effort to analyse and consider user preferences over time, it is difficult to know whether a particular system is doing its task well or altering user preferences to make its task easier. Systems that learn user preferences are at best likely to impact them during the process, and at worst likely to manipulate them to suit their own objective function in a process called *Autoinduced Distributional Shift*. Moreover, 'undesirable behaviours that can arise despite a system's formal correctness'[117]. Moreover, as known from studies on the UCPD and the GDPR, **intent will often be difficult to prove and to bring in connection with a specific outcome in AI settings**[118].

Lastly, we would like to point to the scope of Art. 5(1)(d) AIA, which by introducing a distinction between public and private use of real-time BIS (relying among other heavily on the capturing and processing of biometric data) diverges from the rationale of the GDPR, which does not distinguish between public and private data controllers. This difference in scope is significant, as the GDPR was designed to take into account the overlap of public and private data applications and to empower data subjects with equally applicable opt-out rights. The fact that the AIA – a general regulation itself and a pillar to the EU's digital plan – questions the harmonisation purpose and bearings of the AIA.

6.2.2. Procedure

We next turn our attention to the divergences that relate to procedural issues. We find these, for instance, in the interplay between Art. 10 AIA and Art. 6 and 7 draft ePrivacy Regulation (*lex specialis*

[116] Hildebrandt, M. The Proposal for an EU AI Act of 21 April 2021. Brief Commentary. 19 July 2021, p. 3.

[117] Russell, S., Dewey, D., and Tegmark, M., 'Research Priorities for Robust and Beneficial Artificial Intelligence', *AI Magazine*, 36(4), 2015.

[118] Hacker, 2021, p. 27.

to the GDPR) described earlier, or in the interplay of Art. 10 AIA and the GDPR in general for social scoring applications deployed by the private sector.

Art. 10 AIA reflects the fundamental importance of training data for AI systems that are developed on the basis of supervised learning and reinforcement learning, which in turn form the basis for most AI applications, including facial recognition and scoring ones[119]. The AIA itself, however, does not go beyond broad data quality requirements. It refers instead in its Recital to the general applicability of the GDPR, creating thereby **a number of issues when tracing the data protection rules that supplement Art. 10 AIA**. To begin with, when the GDPR refers to 'the user' it means the data subject, while under the AIA the user of an AI system is actually the GDPR's data controller. It is further unclear whether AI systems qualifying as high-risk under the AIA trigger an assumption of high-risk data processing under Art. 35 GDPR and what that entails[120]. To make things more convoluted, the draft ePrivacy Regulation harbours specific rules for the handling of data obtained from electronic communications and lists purpose exceptions to the strict consent requirements. The purpose of training, however, is not part of the Regulation's consent exception. Against this background, **the proper procedures around training data for AI systems, including such for obtaining consent and from whom, as well as the legal basis for such processing, are unclear**. The individual (data subject) can thus hardly make use of protection mechanisms envisioned in such cases – rights of restriction of processing (Art. 18 and 20 GDPR) as well as rights to deletion and erasure of data (Art. 16 and 19 GDPR).

In terms of procedure, the interplay between the rest of the provisions on private sector social scoring (Art. 6 ff. AIA) and Art. 22 GDPR is also ambiguous. As the AIA itself does not contain individual rights, it is questionable to what extent redress can be sought by means of Art. 22 GDPR and its prohibition on automated decision-making (prohibition of automated data processing to evaluate certain individual traits or characteristics).

[119] Hacker, P. 'A Legal Framework for AI Training Data.' *Law, Innovation and Technology* (forthcoming), 2020.

[120] Codagnone, C. et Al., Identification and assessment of existing and draft EU legislation in the digital field, Study for the special committee on Artificial Intelligence in a Digital Age (AIDA), Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg. 2022, p. 63.

7. Policy Options

7.1. Policy Option 1: Address the incoherence of risk assessment and introduce explicit risk criteria

The AIA is different from other recent EU legislative endeavours such as the GDPR in the sense that the normal risk-based approach to regulating a technology has been expanded from being a procedural obligation to also add a substantial part. The act proposes a classification of risks in relation to AI by dividing applications of AI in 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 AIA's novel element – the combination of a risk-based approach (which is intended to make technology-neutral legislation) with a technology-specific risk classification regime – can potentially 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 other are put in a category based on a self-performed risk assessment (and are categorised via a risk-based approach). When looking at the list, there seems to be an indirect assumption that the more complex the AI system is, the greater the risk.

We have seen in many cases of actualised harm that this is not at all the case. Such assumptions about the risk of applications of a particular technology stem for this element of stating something technology-specific in a regulation: **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. As many risks of AI causing harm are indirect (as AI-based systems will most likely sit in other systems rendering the cut-off value when deciding on the unit of risk analysis very difficult) and will in our application areas of biometrics, social scoring and manipulative systems most likely manifest themselves over time even if the system is legally or formally correct[121], it is difficult to see how an ex-ante risk assessment obligation will be adequate in anticipating risks of AI systems beforehand, let alone if the risks go beyond financial risks or reputational risk (e.g. those risks that private actors would want to focus on primarily).

As our analysis has shown, **the 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 above 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 the two – a policy option would be **to apply risk categories and the obligations they bring about more coherently within the 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 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

[121] Russell, S., Dewey, D., and Tegmark, M., 'Research Priorities for Robust and Beneficial Artificial Intelligence', *AI Magazine*, 36(4),2015, call this a validity problem; 'validity is concerned with undesirable behaviours that can arise despite a system's formal correctness'.

in democratic societies for public institutions offer more fail-safe mechanisms than forms of self-regulation delegated to the private sector. The AIA is not very clear on these considerations.

When looking at external diverging effects related to risk assessment, a policy option would be **to make very clear what the risk assessment is precisely about and to provide clear delineations or cut-off points**. We cannot separate the risks of algorithms and models 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, we caution that the AIA can potentially introduce contradictory risk assessments and a confusion about which risk assessment precedes or weighs more than others in a particular technology's development or deployment. Think for, instance, of overlaps and divergences between the GDPR, which via Article 22 regulates automated processing and profiling (so also potentially forms of AI) and the AI act. **Providing guidelines on how the AIA risk assessment interacts with other risk assessment obligations** put forward in many of the EU regulations and directives that deal with digitisation (e.g. the DPIA in the GDPR) would be a step towards regulatory coherence.

7.2. Policy Option 2: Consider strengthening information and disclosure obligations with withdrawal rights

The 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. A 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, discussed earlier. 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 (basing himself 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 are still rare. Transparency in the form of an AI-registry, as some municipalities in Europe are doing[122], 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 only for research purposes – social media and other big digital service providers aiming to protect their IP. Even if individual harm resulting from an AI system would be 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[123], 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 AIA,**

[122] See for instance Amsterdam Algoritmeregister

[123] Note that consent is still handled from a data processing perspective, and not from such as 'being part of an AI-based prediction system', even though the term 'processing' could cover a wide range of data manipulation techniques.

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 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 AIA would actually decrease trustworthiness. Note here again the difference to the GDPR, which imposes numerous transparency and information obligations to 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 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 unsatisfactory state within the AIA would be: 1) to clarify and directly stipulate in the AIA's provisions **how GDPR rights and remedies are applicable to the addressees of AI systems**, especially so when data rights are involved; and 2) to further critically assess the connection between the AIA's transparency obligations and redress mechanisms by **strengthening information and disclosure obligations with withdrawal rights**. The latter would significantly remedy individuals' legal standing, and would contribute in particular to balancing out the private sector's asymmetrical AI powers.

7.3. Policy Option 3: Consider co-regulation strategies and impact assessments

AI systems are both transformative and disruptive, and their impact on governments, companies and citizens is growing as increasingly complex. However, as documented by the present study, such impact is neither equally distributed nor accounted for. **The AIA appears in some instances to be regulatory strong-arming with 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[124].

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**[125]. 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 AIA that places them in a more prominent perspective**.

The AIA's current approach of governing is such of hierarchy (law-centric) combined with self-regulation (techno-centric). Obligations that are not clarified in a top-down manner are left to the industry standardisation bodies to figure out, which severs the channels of communication between industry and external stakeholders. More importantly, this approach focuses largely on the material

---

[124] Gritsenko, D., and Wood, M., 'Algorithmic governance: A modes of governance approach.' *Regulation & Governance*, 2020.

[125] Cobbe, J., and Singh, J. 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges.' *Forthcoming in Computer Law & Security Review*, 2021.

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 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[126].

The combination of a technology-centric approach with law centric one is something that needs monitoring if and how, that is to say in which sectors and via which obligations and oversight bodies, it will play out. The AIA does not provide clear measures in place to monitor such regulatory effects, except the ex-ante risk assessment and perhaps 'by-design' approaches via 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 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.

[126] Norman, D. A. The Design of Everyday Things. Revised and expanded edition. New York, New York: Basic Books, 2013.

8. References

Agarwal, S., Alok, S., Ghosh, P. & Gupta, S., Financial Inclusion and Alternate Credit Scoring: Role of Big Data and Machine Learning in Fintech, *Indian School of Business*. SSRN, 2021.

Algorithm watch. Automating Society Report 2020.

Algorithm Watch. Denmark - Automating Society Report 2020.

Algorithm Watch. France - Automating Society Report 2020.

Algorithm Watch. Personal Scoring in the EU: Not quite Black Mirror yet, at least if you're rich, 2019.

Algorithm Watch. Spain - Automating Society Report 2020.

Algorithm Watch, New report highlights the risks of AI on fundamental rights, 18 December 2020.

ANWB, ANWB Veilig Rijden Autoverzekering, ANWB website (in Dutch).

Amnesty International, Ban dangerous facial recognition technology that amplifies racist policing, Amnesty International, 2021.

ASR, Kom in beweging met a.s.r. Vitality, ASR website (in Dutch).

BEUC., Artificial Intelligence what consumers say, 2020.

Behavioral Signals, AI-Mediated Conversations, n.d.

CEN-CENELEC, CEN-CLC Response to EC White Paper on AI, June 2020.

CNIL, Reconnaissance faciale: la CNIL met en demeure CLEARVIEW AI de cesser la réutilisation de photographies accessibles sur internet, CNIL, 16 December 2021

Cobbe, J., and Singh, J. 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges.' *Forthcoming in Computer Law & Security Review*, 2021.

Codagnone, C. et Al., Identification and assessment of existing and draft EU legislation in the digital field, Study for the special committee on Artificial Intelligence in a Digital Age (AIDA), Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg. 2022.

Cognitech, Commercial/Service-Based Applications, Cognitech website.

Credolab, Digital Credit Scoring with Alternative Data, Credolab website.

Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC, OJ L 218/82.

Delacroix, S., 'Beware of 'Algorithmic Regulation'', *SSRN*, 2019.

Delipetrev, B., Tsinaraki, C., Nepelski, D., Gomez Gutierrez, E., Martinez Plumed, F., Misuraca, G., De Prato, G., Fullerton, K., & Craglia, M., and Duch-Bro, N., 2020. 'AI Watch 2019 Activity Report,' JRC Working Papers JRC121011, Joint Research Centre (Seville site).

Descript, Overdub: Ultra realistic text to speech voice cloning.

Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business to consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), OJ L 149/22, art 5.

Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119/89 ('Law Enforcement Directive').

Dumbrava, C., 'Artificial Intelligence at EU borders - overview of applications and key issues', EPRS, European Parliament, July 2021.

Ebers. M., Hoch, V., Rosenkranz, F., Ruschemeier, H. and Steinrötter, B., The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS), *J Multidisciplinary Scientific Journal*, 4(4), 2021, p. 592.

European Association for Digital Transition, Big tech lobbying in Brussels: A lot of money. Too much influence?, 4 March 2021.

European Commission, Data Act: Commission proposes measures for a fair and innovative data economy, press release, European Commission, 23 February 2022.

European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Fostering a European Approach to Artificial Intelligence (COM(3032) 205 Final)' (21 April 2021) p. 2.

European Commission/European Parliament, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021) 206 final).

European Commission/European Parliament, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM(2020) 825 final).

European Commission/European Parliament, Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) (COM(2020) 842 final).

European Commission/European Parliament, Proposal for a Regulation of the European Parliament and of the Council on machinery products (COM(2021) 202 final) (Machinery Regulation).

European Commission/European Parliament, Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) (COM(2020) 767 final).

European Commission / European Parliament, Proposal for a regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) (COM(2017) 10 Final).

European Commission/European Parliament, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

European Commission, European Data governance act, European Commission website.

European Commission, Proposal for a Regulation on Privacy and Electronic Communications, European Commission website.

European Commission, Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64, European Commission, 19 February 2020.

European Commission, Regulatory framework on AI, European Commission website.

European Commission, The Digital Services Act package, European Commission website.

European Commission, The Digital Services Act: ensuring a safe and accountable online environment, European Commission website.

European Commission, A Europe fit for the digital age, European Commission website.

European Commission, Shaping Europe's digital future, European Commission, February 2020.

European Data Protection Board, Dutch DPA issues Formal Warning to a Supermarket for its use of Facial Recognition Technology. European Data Protection Board website, 2021.

European Data Protection Board, Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications, 28 January 2020.

European Parliament, AI rules: what the European Parliament wants, European Parliament website.

European Parliament/ Special Committee on Artificial Intelligence in a Digital Age, Draft report on artificial intelligence in a digital age (2020/2266(INI), PR_INI (europa.eu), p. 8.

EUobserver., 'Deepfakes' - a political problem already hitting the EU, 2021.

Financial Times, EU targets Big Tech with 'hit list' facing tougher rules, 11 October 2020.

Floridi, L., 'The European Legislation on AI: a Brief Analysis of its Philosophical Approach.' *Philosophy & Technology* (2021), pp. 1-8.

Fölster, M., Hess, U., & Werheid, K. 'Facial age affects emotional expression decoding'. Front. Physchol, 2014.

Funding Matters, Statements about the corporate funding of academia, Funding Matters website.

Galson, W., 'Is seeing still believing? The deepfake challenge to truth in politics', The Brookings Instutition., January 2020.

Gemeente Amsterdam, Amsterdam Algoritmeregister, Gemeente Amsterdam, Amsterdam Algoritmeregister, website of the Municipality of Amsterdam, 2020.

Gonzalez Fuster, G. and Nadolna Peeters, M., Person Identification, Human Rights and Ethical Principles: Rethinking biometrics in the era of Artificial Intelligence, EPRS, European Parliament, December 2021

Gritsenko, D., and Wood, M., 'Algorithmic governance: A modes of governance approach.' *Regulation & Governance*, 2020.

Guihot, M. 'Coherence in technology law.' *Law, Innovation and Technology*, 11(2), 2019, pp. 311-342.

Hacker, P. 'A Legal Framework for AI Training Data.' *Law, Innovation and Technology (forthcoming)*, 2020.

Hacker, P. Manipulation by algorithms. Exploring the traingle of unfair commercial practice, data protection and privacy law, *European Law Journal*, 2021, p. 28.

Hildebrandt, M. The Proposal for an EU AI Act of 21 April 2021. Brief Commentary, 19 July 2021.

Huijstee, M. et al., Tackling Deepfakes in European Policy, EPRS, European Parliament, July 2021.

Internet Policy Review, Identifying harm in manipulative artificial intelligence practices, Internet Policy Review, 30 November 2021.

Interesting engineering, AI Facial Recognition and IP Surveillance for Smart Retail, Banking, and the Enterprise, 27 January, 2020.

Kleinig, J., 'Crime and the Concept of Harm.' *American Philosophical Quarterly*, 15(1), 1978, pp. 27-36.

Koops, B. 'Ten Dimensions of Technology Regulation-Finding Your Bearings in the Research Space of an Emerging Discipline.' *Dimensions of technology regulation*, 2010, p. 319.

Kosinski, M., Facial recognition technology can expose political orientation from naturalistic facial images. *Scientific reports*, 11(1), 2021, 1-7.

Mahler, T., 'Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal.' *Nordic Yearbook of Law and Informatics*, 2021.

Norman, D. A. The Design of Everyday Things. Revised and expanded edition. New York, New York: Basic Books, 2013.

NOS. Veel onduidelijk over 'deepfake-gesprek' van Kamerleden met medewerker Navalny, 24 April 2021.

OECD. 'Personal Data Use in Financial Services and the Role of Financial Education' 2020.

Office of the Privacy Commissioner of Canada, Announcement: Clearview AI ordered to comply with recommendations to stop collecting, sharing images, Office of the Privacy Commissioner of Canada website, 2021.

Rechtbank Den Haag, ECLI:NL:RBDHA:2020:1878, SyRI court ruling. Rechtbank Den Haag. 6 March 2020.

Rechtspraak.nl, SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens, rechtspraak.nl.

Realeyes, About us - Realeyes, Realeyes website.

Reuters, Rohingya refugees sue Facebook for $150 billion over Myanmar violence, December 2021.

Rowe, E. A., Regulating Facial Recognition Technology in the Private Sector. *Stanford Technology Law Review*, 24(1), 2020.

Russell, S., Dewey, D., and Tegmark, M., 'Research Priorities for Robust and Beneficial Artificial Intelligence', *AI Magazine*, 36(4), 2015.

TALA, Data Ethics, Tala website.

Taylor, L. Public actors without public values: legitimacy, domination and the regulation of the technology sector. *Philosophy & technology*, 34(4), 2021, pp.897-922.

TechCrunch, Clearview AI told to stop processing UK data as ICO warns of possible fine, 29 November 2021.

TechCrunch, France latest to slap Clearview AI with order to delete data, 16 December 2021.

Towards Data Science, 7 Types of AI Risk and How to Mitigate their Impact, September 2020.

The Guardian. European MPs targeted by deepfake video calls imitating Russian opposition, 22 April 2021.

The Guardian, Opinion: Billionaire capitalists are designing humanity's future. Don't let them, February 2021.

The Guardian, Who won and who lost: when A-levels meet the algorithm, 13 August 2020.

The Markup., Algorithms Behaving Badly: 2020 Edition, December 2020.

The New York Times, Cambridge Analytica and Facebook: The Scandal and the Fallout So Far, April 2018.

The Washington Post, Another fake video of Nancy Pelosi goes viral on Facebook, 3 August 2020.

The Washington Post, Civil rights groups push Biden administration to take stand against facial recognition, 17 February 2021.

Thales, Smart gates for Paris Orly & Charles de Gaulle, Thales website.

Schiphol, Facial recognition pilot for departing travellers, Schiphol website.

Schiphol, Travel with facial recognition, Schiphol website.

Smuha, N., Ahmed-Rengers, E., Harkens, A., Li., W., MacLaren, J., Piselli, R. and Yeung, K., How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act, SSRN, 2021.

Youtube, Human-centred AI in the EU, YouTube for more elaboration on risks and potential harms, 30 November 2020.

Van Dijk, N., Gellert, R., & Rommetveit, K., 'A risk to a right? Beyond data protection risk assessments.' *Computer Law & Security Review*, 32(2), 2016, pp. 286-306.

Van Hoecke, M. and Warrington, M., 'Legal cultures, legal paradigms and legal doctrine: towards a new model for comparative law.' *International & Comparative Law Quarterly*, 47(3), 1998, pp. 495-536.

Veale, M., and Zuiderveen Borgesius. F., 'Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach.' *Computer Law Review International* 22(.4), 2021, pp. 97-112.

VentureBeat, Realeyes raises $12.4 million to help brands detect emotion using AI on facial expressions, 6 June 2019.

Vice, Car Companies Want to Monitor Your Every Move With Emotion-Detecting AI, 27 July 2020 and The Sun, BMW's new X5 SUV has a camera that alerts you when your eyes wander off the road, 2 October 2018.

WeSee, Emotion Recognition Could Eliminate Insurance Fraud, WeSee website.

Wired, The Dark Side of Big Tech's Funding for AI Research, Wired.

Wired, The Dark Side of Big Tech's Funding for AI Research, 10 December 2020.

Wired, Job Screening Service Halts Facial Analysis of Applicants, 12 January 2021.

Wired, Face Recognition Is Being Banned—but It's Still Everywhere, 22 December 2021.

World Bank, The unbanked - Global Findex, World Bank website.

This study identifies and examines sources of regulatory divergence within the AI act regarding the obligations and limitations upon public and private sector actors when using certain AI systems. A reflection upon possible impacts and consequences is provided, and a range of policy options is suggested for the European Parliament that could respond to the identified sources of divergence.

The study is specifically focused on three AI application areas: manipulative AI, social scoring and biometric AI systems. Questions regarding how and when those systems are designated as prohibited or high-risk and the potentially diverging obligations towards public versus private sector actors and the rationale behind it, are described.

QA-07-22-331-EN-N

This is a publication of the Scientific Foresight Unit (STOA)
EPRS | European Parliamentary Research Service

ISBN 978-92-846-9459-4 | doi:10.2861/69586 | QA-07-22-331-EN-N