HOW THE EU CAN ACHIEVE LEGALLY TRUSTWORTHY AI:

A RESPONSE TO THE EUROPEAN COMMISSION'S PROPOSAL FOR AN ARTIFICIAL INTELLIGENCE ACT

by Nathalie Smuha,[a] *Emma Ahmed-Rengers*[b]
Adam Harkens,[c] *Wenlong Li,*[d] *James MacLaren,*[e] *Riccardo Piselli*[f]
and Karen Yeung[g]

LEADS Lab @University of Birmingham
For a Legal, Ethical & Accountable Digital Society

5 August 2021

[a] Researcher and FWO Scholar, KU Leuven Faculty of Law, Department of International and European Law. Visiting Researcher, University of Birmingham School of Law, LEADS Lab.
[b] Doctoral Researcher, University of Birmingham School of Law & School of Computer Science, LEADS Lab.
[c] Postdoctoral Researcher, University of Birmingham School of Law, LEADS Lab.
[d] Postdoctoral Researcher, University of Birmingham School of Law, LEADS Lab.
[e] Doctoral Researcher, University of Birmingham School of Law, LEADS Lab.
[f] Postdoctoral Researcher, University of Birmingham School of Law, LEADS Lab.
[g] Professor of Law, Ethics and Informatics, University of Birmingham School of Law & School of Computer Science, Head of the LEADS Lab.

EXECUTIVE SUMMARY

This document contains the response to the European Commission's Proposal for an Artificial Intelligence Act laying down harmonised rules for Artificial Intelligence (AI) from members of the Legal, Ethical & Accountable Digital Society (LEADS) Lab at the University of Birmingham.

The Proposal seeks to give expression to the concept of 'Lawful AI' which was not addressed by the *Ethics Guidelines for Trustworthy AI* published in 2019 by the Commission's High-Level Expert Group on AI, as it confined its discussion to the concepts of 'Ethical' and 'Robust' AI. We develop the concept of 'Legally Trustworthy AI,' arguing that it should be grounded in respect for three pillars on which contemporary liberal democratic societies are founded, namely: fundamental rights, the rule of law, and democracy. To promote Legally Trustworthy AI, the Proposal must ensure (1) an appropriate allocation and distribution of responsibility for the wrongs and harms of AI; (2) a legitimate and effective enforcement architecture, including adequate mechanisms for transparency to secure the effective protection of fundamental rights and the rule of law, and to provide clear, stable guidance to legal subjects in a manner that coheres with other applicable laws; (3) adequate rights of public participation, and information rights necessary to ensure meaningful accountability for the development, deployment and oversight of AI systems that is necessary in a democratic society.

We welcome many aspects of the Proposal, including its commitment to dealing with the risks of AI through a set of obligations and a public enforcement mechanism; the more refined nature of its risk-based classification of AI systems compared to the Commission's White Paper on AI of February 2020; the introduction of prohibited practices; and the introduction of a European database and logging requirements for high-risk AI systems. We wholeheartedly support the Proposal's aim to protect fundamental rights. However, as currently drafted, the Proposal does not provide adequate fundamental rights protection, nor does it provide sufficient protection to maintain the rule of law and democracy, and thus fails to secure Legally Trustworthy AI. Accordingly, the Proposal suffers from the following three problems:

Firstly, it fails to reflect fundamental rights as claims with enhanced moral and legal status which subjects any rights interventions to a demanding regime of scrutiny that must satisfy tests of necessity and proportionality for limited and narrowly defined purposes. Additionally, the Proposal does not always accurately recognise the wrongs and harms associated with different kinds of AI systems and appropriately allocate responsibility for harms and wrongs.

- This problem is reflected in the lack of clarity concerning its scope. Hence, we recommend revision and refinement of the definition of AI, the position of academic AI researchers, the status of military AI, and the role of national security agencies. This problem is also reflected in the incomplete list of prohibited practices, and the failure to include mechanisms for its review and revision.
- The Proposal should include stronger protection against AI-enabled manipulation, social scoring and the use of biometric identification systems. Moreover, it should prohibit the use of emotion recognition systems and the use of live remote biometric categorisation systems by law enforcement and by other public actors with coercive powers, or private actors acting on their behalf. Given their intrusive nature, it is not understandable why emotion recognition systems and biometric categorisation systems were not at the very least recognised as high-risk.
- The provisions for high-risk AI systems do not provide sufficient protection against the actual and threatened harms and wrongs generated by AI, including fundamental rights violations. They leave too much discretion to AI providers to decide on politically

charged matters, which stems from the Proposal's poor understanding of fundamental rights. Moreover, the eight-point list of high-risk systems must be amendable to ensure that it is future-proof and fit for purpose. Finally, the requirements for the data governance, transparency, and human oversight of high-risk systems would benefit from clarification and improvement.

- The proposal does not appear to ensure that fundamental rights interferences by AI are subjected to demanding tests of necessity and proportionality for narrowly defined purposes, nor to consistently allocate legal responsibility for the wrongs and harms of AI in an appropriate manner, thus failing to secure the first pillar of Legally Trustworthy AI.

Secondly, the Proposal does not ensure an effective framework for the enforcement of legal rights and duties.

- It envisages an inappropriately large role for AI providers in the implementation of the Regulation, granting them too much discretion and putting undue faith in the effectiveness of conformity assessment and CE marking.
- In doing so, it risks creating incoherence between the proposed Regulation and other EU legal instruments, such as fundamental rights law, the GDPR, the Law Enforcement Directive and MiFID II.
- The Proposal also fails to recognise the status of individuals adversely affected by AI systems in its enforcement mechanisms, reflected in a complete lack of procedural rights for individuals, such as the rights to contest and to seek redress, and the lack of an adequate complaints mechanism.
- Finally, the Proposal's enforcement mechanism relies heavily on national competencies without providing assurances regarding the proper staffing and funding of the Member State authorities in charge of enforcing the proposed Regulation. As the Proposal's enforcement architecture has significant weaknesses and its coherence with other EU legal instruments is currently wanting, the Proposal fails to adequately protect the rule of law, as the second pillar of Legally Trustworthy AI.

Thirdly, the Proposal neglects to ensure meaningful transparency, accountability, and rights of public participation, thereby failing to provide adequate protection for democracy as the third pillar of Legally Trustworthy AI. In particular:

- The public is not provided with consultation and participation rights regarding future revisions of the list of high-risk AI systems, nor regarding the determination of what constitutes an 'acceptable' residual risk in the context of high-risk AI systems.
- The Proposal does not provide individuals with substantive rights not to be subjected to prohibited or otherwise noncompliant AI systems, illustrating the Proposal's complete lack of attention to 'ordinary people.' Nor are individuals granted meaningful information rights to enable them to form informed opinions and contest the development and deployment of controversial AI systems.
- The Proposal does not provide for democratic input on the development of the technical standards crucial for the implementation of the proposed regulatory framework.

Based on the three pillars of Legally Trustworthy AI and the abovementioned concerns, we make detailed recommendations for revisions to the Proposal, which are listed in the final Chapter of this document.

TABLE OF CONTENTS

1. INTRODUCTION

This document sets out the response to the European Commission's Proposal for an Artificial Intelligence Act laying down harmonised rules for Artificial Intelligence (AI)[1] from members of the Legal, Ethical & Accountable Digital Society (LEADS) Lab at the University of Birmingham.

We are grateful for the opportunity to voice our views on this unique and ground-breaking piece of legislation, which we very warmly welcome. We wholeheartedly support the Commission's commitment to providing legally binding norms and standards for the development and use of AI, as we believe that it is high time for a robust and coherent framework which prevents the gravest harms and wrongs of AI and allocates responsibility for the risks associated with this exciting yet powerful and intrusive technology. Throughout this document, we stress the importance of the legal nature of the Proposal, which sets it apart from all other initiatives on AI ethics or technical best practices. In this context, we expressly welcome the Proposal's focus on protecting fundamental rights – which are not only moral precepts, but also legal rights that include a concrete and well-established framework for interpretation, implementation and enforcement. Yet, despite the Proposal's significant achievement of providing legally binding rules for AI, our analysis shows that the Proposal does not yet live up to its promise. It suffers from a number of deficiencies which compromise the value and importance of its legal nature, and which need to be addressed in order to ensure that the Proposal can fulfil its noble aspirations.

To explain these shortcomings, we sketch a tripartite analytical framework based on the concept of Legally Trustworthy AI. We also offer practical suggestions on how these shortcomings might be addressed. Our evaluation of the Proposal does not adopt the perspective of EU constitutional law. Nor does it offer a doctrinal legal analysis which evaluates the Proposal against the internal requirements of EU law. Although we recognise that not all amendments we suggest might be possible considering the current legal basis of the Proposal and the competences of the different bodies of the EU legal order, we urge the Commission to take seriously our recommendations – if not in this Proposal, then in another legal instrument.

Our analysis begins with the earlier work done towards establishing Trustworthy AI in Europe by the European Commission and its High-Level Expert Group on AI (HLEG), which provides the background against which the Proposal must be understood. In their *Ethics Guidelines for Trustworthy AI*,[2] the HLEG provided the contours for 'Trustworthy AI' – defined as AI which is lawful, ethical and robust. This Proposal represents an important attempt to address the requirement of 'Lawful AI' which was excluded from the HLEG's remit. On the basis of this requirement of Lawful AI, we develop the concept of 'Legal Trustworthiness,' which we argue rests on three pillars upon which contemporary liberal democratic societies are rooted: ensuring respect for fundamental rights, the rule of law, and democracy. The three pillars of Legally Trustworthy AI in turn require that the regulation of AI: (1) appropriately allocates responsibility for the harms and wrongs resulting from AI systems, especially where these pertain to fundamental rights; (2) establishes and maintains a coherent legal framework

[1] 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," *EUR-Lex*, April 21, 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206. Hereafter referred to as "the Proposal," "the proposed Regulation," or the "draft Regulation."

[2] High-Level Expert Group on AI, "Ethics Guidelines for Trustworthy AI," 8 April 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

accompanied by effective and legitimate enforcement mechanisms to secure and uphold the rule of law; and (3) places democratic deliberation at its centre, which includes the conferral of public participation and other information rights necessary for effective democracy.

In what follows, we first outline the welcome aspects of the proposed Regulation (Chapter 2). Then, we explain why Legally Trustworthy AI should be the core mission of the Proposal, and how Legal Trustworthiness can be attained through respect for each of its three pillars (Chapter 3). We argue that the Proposal falls short of each of these three pillars (Chapter 4), setting out our concrete recommendations (Chapter 5). In particular, we recommend significant changes concerning the Proposal's understanding of fundamental rights; its scope; its provisions for prohibited systems, biometric systems and high-risk systems; its enforcement mechanisms; the (lack of a) role allocated to the individuals subjected to or otherwise directly affected by AI systems; and the lack of meaningful public participation and information rights.

2. WELCOME ASPECTS OF THE PROPOSED REGULATION

Before we address the ways in which we believe the proposed Regulation could be improved, we emphasise that we wholeheartedly welcome the European Commission's initiative to introduce legally enforceable standards for the development and use of AI. With its ground-breaking Proposal, the Commission explicitly acknowledges both the adverse effects on fundamental rights and safety that AI can generate, and that voluntary self-regulation offers inadequate protection.

By establishing a public enforcement structure, the Commission affirms that the risks associated with AI need to be dealt with as a matter of public interest, and in a manner that secures a level playing field of protection across the EU, regardless of an individual's Member State of residence. Despite being primarily based on a legal basis meant to harmonise the internal market, the Proposal seeks to affirm and operationalise a shared commitment to upholding fundamental rights.

While various aspects of the Proposal merit significant improvement, the Proposal presents a concrete and systematic way forward to regulate AI, which in turn provides EU legislators and stakeholders with an opportunity to discuss and refine this approach before adoption. The Proposal's underlying ambition of prioritising respect for fundamental rights (including safety and health) is both important and warmly welcomed (although, as further detailed below, we believe this foundational commitment must be accompanied by the protection of democracy and the rule of law, and requires a strengthening of the Proposal in order to be achieved). With this initiative, the EU sends a strong signal that it takes the risks and dangers raised by AI seriously, and does not shy away from taking legally-enforceable measures – including the imposition of prohibitions or 'red lines.' We hope that this ambition will act as a catalyst for regulatory action in this field by other legislators across the world.

Compared to the blueprint for a potential AI regulation set out in White Paper on Artificial Intelligence (February 2020), this Proposal shows significant improvement in several respects.

First, the Proposal takes a more nuanced stance towards the 'risk-based' classification of AI systems, swapping the binary risk-categorisation (high-risk/no-risk) for a more refined approach. Besides distinguishing between (1) AI practices that pose an unacceptable risk and are hence prohibited (Title II), (2) AI practices which pose a high risk to the fundamental rights or to the safety of natural persons and are subject to specific requirements (Title III), (3) AI practices which require enhanced transparency (Title IV) and (4) AI practices which pose no risk or a low level of risk yet may benefit from a voluntary code of conduct (Title IX), the Proposal also renders explicit the criteria which are considered relevant to determine the 'high-risk' nature of an AI system (in Article 7). Nevertheless, as we explain below, the protection

afforded by the specified risk criteria should be refined and enhanced, and it is questionable whether the listed AI practices are appropriately categorised.

Secondly, we welcome the introduction of prohibited practices as per Title II of the proposed Regulation, which constitutes an important addition to the White Paper of February 2020. By providing a set of 'red lines,' the Commission acknowledges that certain uses of AI are by definition incompatible with fundamental rights. While the scope of these prohibitions needs to be broadened, this acknowledgment is an important step towards securing fundamental rights protection in Europe.

Thirdly, the Proposal introduces a European database, managed by the European Commission, in which essential information on all stand-alone high-risk AI systems deployed in the EU is to be registered. This database could increase urgently needed transparency around the use of potentially problematic AI systems and may be a useful resource for citizens and civil society organisations. Moreover, this database could be a helpful tool for national authorities, allowing them to strengthen public supervision and advancing the proposed Regulation's enforcement. However, we believe the matters which must be disclosed need significant expansion if the level of public transparency provided under the regime is to meet threshold requirements of healthy democratic societies.

Fourthly, the Proposal introduces requirements to ensure that the mitigation of risks arising from problematic AI systems does not remain a dead letter, creating the *possibility* for meaningful implementation and enforcement. The introduction of mandatory logging requirements for high-risk AI systems is, for example, a welcome development in this regard, as is the need to specify the purpose of the AI system in a sufficiently concrete manner to allow for context-specific testing mechanisms. Furthermore, if the user of the AI system subsequently uses it for a different purpose, the 'AI user' (whose main obligation is to ensure human oversight over the AI system once deployed) becomes the *provider* of an AI system, hence incurring the specific obligations imposed on providers rather than users (as set out under Article 28). Finally, the Proposal appears to acknowledge that testing technical aspects of an AI system may lead to different results in a lab environment and in the real-world, and recognises that these differences need to be taken into account when seeking to comply with the Regulation's various testing requirements.

At the same time, the proposed Regulation remains deficient in numerous ways, both in its substance and enforcement architecture. While proclaiming the protection of fundamental rights as a core aim, the Proposal does not appear to take fundamental rights sufficiently seriously. The proposed Regulation also fails to adequately reflect the EU's commitment to *democracy* and the *rule of law*, which – together with respect of *fundamental rights* – can be seen as constituting the three core pillars of the EU legal order.[3]

In what follows, we first outline the contours of the concept of Legally Trustworthy AI with its three pillars of fundamental rights, rule of law, and democracy. We use these pillars as a measure for the quality of the Proposal. We argue that the Proposal falls short of all three of

[3] Article 2 Treaty on European Union sets out the values on which the EU is founded, these being "common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail." On the importance of the triad of human rights, democracy and the rule of law in the context of AI and the EU, see also e.g., Paul Nemitz, "Constitutional Democracy and Technology in the Age of Artificial Intelligence," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2133 (2018). See also HLEG, "Ethics Guidelines," 4: "It is through Trustworthy AI that we, as European citizens, will seek to reap its benefits in a way that is aligned with our foundational values of respect for human rights, democracy and the rule of law."

the pillars of Trustworthy AI, after which we make specific recommendations on how the Proposal should be revised and improved.

3. THE THREE ESSENTIAL FUNCTIONS OF LEGALLY TRUSTWORTHY AI

Our analysis of the proposed Regulation is rooted in the concept of 'Trustworthy AI' set out in two primary publications produced by the European Commission's High-Level Expert Group on AI (hereafter the 'HLEG'): the *Ethics Guidelines for Trustworthy AI*[4] (hereafter the 'Ethics Guidelines') and the *Policy and Investment Recommendations for Trustworthy AI* (hereafter the 'Policy Recommendations').[5] In the following sections, we explain why Trustworthy AI requires 'Legal Trustworthiness,' what 'Legally Trustworthy AI' is, and how it can be attained. The rest of this document then sets out how and where the Proposal falls short of the requirements of Legal Trustworthiness, and therefore risks failing its core mission.

3.1 Achieving 'Legally Trustworthy AI' as the Proposal's core mission

The Proposal's aims and objectives must be understood in the context of earlier work on 'Trustworthy AI' by the HLEG and by the European Commission which endorsed the HLEG's work.[6] The underlying purpose of the Proposal is to fill the gap that the Ethics Guidelines left open regarding 'Lawful' or 'Legally Trustworthy' AI.

The Ethics Guidelines state that, "[i]t is through Trustworthy AI that we, as European citizens, will seek to reap its benefits in a way that is aligned with our foundational values of respect for human rights, democracy and the rule of law."[7] The Ethics Guidelines also explain that:

> Trustworthy AI has three components, which should be met throughout the system's entire life cycle:
>
> (a) it should be **lawful**, complying with all applicable laws and regulations;
> (b) it should be **ethical**, ensuring adherence to ethical principles and values and;
> (c) it should be **robust**, both from a technical and social perspective
>
> since, even with good intentions, AI systems can cause unintentional harm. Each component in itself is necessary but not sufficient for the achievement of Trustworthy AI. Ideally, all three components work in harmony and overlap in their operation. If, in practice, tensions arise between these components, society should endeavour to align them.[8]

The Ethics Guidelines have no binding legal force[9] and deliberately refrain from considering whether additional legal measures would be needed to ensure 'Lawful' AI. Instead, their operative provisions are confined to 'Ethical' and socio-technically 'Robust' AI, while Lawful AI is not discussed. Yet, the Ethics Guidelines explicitly recognise the critical importance of Lawful AI in supporting the other two components of Trustworthy AI (i.e., the ethical and

---

4 HLEG, "Ethics Guidelines."

5 High-Level Expert Group on AI, "Policy and Investment Recommendations for Trustworthy AI," 26 June 2019, https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

6 European Commission, "Building Trust in Human-Centric Artificial Intelligence," 8 April 2019, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58496.

7 HLEG, "Ethics Guidelines," 4. As the Guidelines specify at page 9, Trustworthy AI is explicitly grounded in a deep commitment to three fundamental principles: "We believe in an approach to AI ethics that is based upon the fundamental rights enshrined in the EU Treaties, the Charter of Fundamental Rights of the EU (EU Charter) and international human rights law. Respect for fundamental rights, within a framework of democracy and rule of law, provides the most promising foundations for identifying abstract ethical principles and values which can be operationalised in the context of AI."

8 HLEG, "Ethics Guidelines," 2.

9 The initial call for experts for the HLEG, stating that the HLEG would propose to the Commission 'AI Ethics Guidelines,' is available at: https://digital-strategy.ec.europa.eu/en/news/call-high-level-expert-group-artificial-intelligence.

socio-technical dimensions). This is affirmed and endorsed in the Policy Recommendations stating that:

> On 8 April 2019, we published our Ethics Guidelines which set out three components for Trustworthy AI: (1) lawful AI, (2) ethical AI and (3) robust AI. The Ethics Guidelines only deal with the two latter components, yet the first is equally crucial.[10]

We understand the primary aim of the proposed Regulation as addressing this third component of Trustworthy AI, by establishing a set of legally binding norms and institutional mechanisms necessary to ensure Lawful AI.

While the HLEG did not explain the concept of Lawful AI, it is best illuminated by considering the key distinction between legal standards on the one hand, and voluntary ethical and technical standards on the other hand. The most powerful features of legal standards concern the fact that – unlike ethical or technical standards – they are mandatory and legally binding: promulgated, monitored and enforced in the context of a system of institutions, norms and a professional community which work together to ensure that laws are properly interpreted, effectively complied with and duly enforced.[11] Mature legal systems consist of institutions, procedures and officeholders which are legally empowered to gather information to monitor compliance with legal duties. Where a possible violation of legal requirements is identified, those with standing may initiate enforcement action before an independent tribunal seeking binding legal remedies, which include both orders to bring any legal violations to an end and, if applicable, the imposition of sanctions and orders for compensation. This institutional structure differentiates law from both ethics and technical best practices. The 'Lawful' component of Trustworthy AI gives the concept of Trustworthiness 'teeth.'

However, the mere existence of some 'teeth' does not *by definition* contribute to or ensure Legal Trustworthiness. As is evidenced by the Commission's efforts to address the requirement of Lawful AI with a whole new regulation, the concept of 'Lawful AI' cannot mean that Trustworthy AI simply complies with whichever legal rules happen to exist. For legality to contribute to 'Trustworthiness,' it is crucial that the legal framework itself adequately deals with the risks associated with AI. This desideratum goes far beyond simple legal compliance checks – it requires the existence of a regulatory framework which addresses the foundational values of fundamental rights, the rule of law, and democracy. As the *Policy Recommendations* state:

> Ensuring Trustworthy AI necessitates an appropriate governance and regulatory framework. By appropriate, we mean a framework that promotes socially valuable AI development and deployment, ensures and respects fundamental rights, the rule of law, and democracy while safeguarding individuals and society from unacceptable harm.[12]

Legally Trustworthy AI (as opposed to simple legally compliant AI) operates at two levels. The first level concerns the extent to which the *regulatory framework around AI* promotes the values of fundamental rights, democracy and the rule of law. The second level concerns the way in which *AI systems themselves* affect those three values. For an AI system to be Legally Trustworthy, it must therefore (1) be regulated by a governance framework which promotes fundamental rights, democracy and the rule of law, and (2) not itself be detrimental to fundamental rights, democracy and the rule of law. The second level is conditional on the first level: if the regulatory framework adequately protects fundamental rights, the rule of law and

---

10 HLEG, "Policy and Investment Recommendations," 37.

11 See generally Peter Cane, *Responsibility in Law and Morality* (Oxford: Hart Publishing, 2002).

12 HLEG, "Policy and Investment Recommendations," 37.

democracy, it does not allow systems which negatively affect fundamental rights, democracy and the rule of law to exist without due precautions and protections.

While the Proposal explicitly claims to protect fundamental rights in the context of AI, we argue that it does not currently provide adequate fundamental rights protection. Moreover, it does not explicitly address the other two pillars upon which liberal constitutional political communities rest and upon which the EU legal order is founded – and explicitly reaffirmed in the Ethics Guidelines – namely that of democracy and the rule of law. Before we outline that the Proposal therefore cannot guarantee Legally Trustworthy AI, we explain how Legally Trustworthy AI is attained.

3.2 What does Legally Trustworthy AI require?

In what follows, we briefly outline three core functions which the legal system must provide to establish Lawful or Legally Trustworthy AI. It builds on the tripartite framework for Trustworthy AI set out in the *Ethics Guidelines*, explained in the previous section. To ensure Legally Trustworthy AI within liberal democratic constitutional systems, the legal system must:

(a) Allocate and distribute responsibilities for wrongs and harms appropriately, which includes the effective protection of *fundamental rights*;

(b) Establish and maintain an effective framework to enforce legal rights and responsibilities, and secure the clarity and coherence of the law itself – these being essential elements of the *rule of law;*

(c) Reflect a commitment to *democracy* by securing meaningful transparency, accountability and rights of public participation to its members.

Accordingly, our analysis and recommendations are rooted in our belief that law must ensure these three core functions are properly established and maintained. Legally Trustworthy AI therefore requires both a regulatory framework for AI which embodies these three functions, and guarantees that the AI systems permitted under the regulatory framework do not undermine these three functions. Each of the dimensions of Legal Trustworthiness - which are closely entwined - are briefly outlined below:

a) Allocate and distribute responsibility for wrongs and harms appropriately, and in a manner that adequately protects *fundamental rights*

Firstly, Legal Trustworthiness requires the appropriate allocation of responsibility for harms and wrongs. A core function of modern legal systems is to provide a binding framework to enable peaceful social cooperation between strangers. The legal system achieves this *inter alia* by attributing legal responsibility to those whose activities produce ‘other-regarding’ harms or wrongs, whether intentional or otherwise, resulting in the imposition of either (or both) civil or criminal liability as appropriate. In this way, the law seeks to reduce and prevent harm to others, and to ensure appropriate redress where such adverse events occur. The law establishes and publicly promulgates legally binding rules which identify the scope and content of the rights and responsibilities of legal and other persons, thereby providing guidance to legal subjects so that they can alter their behaviour accordingly so as not to fall foul of the law’s demands. This legal guidance function plays an important role in protecting the legal rights, interests and expectations of all members of the community against unlawful interference by others.

To this end, *fundamental rights* can often be understood as providing the *justification* for concrete legal standards, instruments and mechanisms, thereby providing legal subjects with

clearer and enforceable forms of legal protection. While fundamental rights can be understood as moral rights, bestowed on individuals by virtue of their moral status as human beings, they are realised in concrete legal arrangements.[13] For example, the General Data Protection Regulation ('GDPR') is known to concretise the fundamental right to personal data protection.[14] Similarly, laws concerning the proper conduct of elections can be understood as rooted in a concern to protect the fundamental right to vote and the right to free and fair elections.

The responsibilities of those involved in designing, testing and putting into service AI systems is of particular importance, as those who are subjected to the adverse effects of those systems might not always be aware of this fact. This is because AI systems can operate in ways that are opaque or even invisible, in real time and at scale. Consider the example of Clearview AI, a US-based company with a database of three billion facial images scrapped from Facebook, YouTube, Venmo and millions of other websites,[15] which has been used by a number of law enforcement agencies and other European actors. Today, almost every law enforcement agency in Europe uses AI-enabled biometric identification technology.[16] Examples are, of course, not limited to the law enforcement sphere. In fact, the list of high-risk AI systems in the Proposal's Annex III provides a good overview of problematic AI-enabled practices, from workplace surveillance to the detection of the emotional states of persons in the context of a migration management.[17]

Legally Trustworthy AI therefore requires a regulatory framework which prevents the gravest harms and wrongs generated by AI systems, and appropriately allocates responsibility for them if they do occur, particularly when they violate fundamental rights.

b) Establish and maintain an effective framework to enforce legal rights and responsibilities, and secure the clarity and coherence of the law itself *(rule of law)*

Secondly, Legal Trustworthiness must ensure adherence to the essential elements of the rule of law including, *inter alia*, effective enforcement, judicial protection, and the coherence of the law.

---

[13] See HLEG, "Ethics Guidelines," 10: "Understood as legally enforceable rights, fundamental rights therefore fall under the first component of Trustworthy AI (lawful AI), which safeguards compliance with the law. Understood as the rights of everyone, rooted in the inherent moral status of human beings, they also underpin the second component of Trustworthy AI (ethical AI), dealing with ethical norms that are not necessarily legally binding yet crucial to ensure trustworthiness." Accordingly, although the Ethics Guidelines recognise that there is often overlap between the requirements of ethics and of law, the two are not coterminous.

[14] On the need for the concretisation of human rights in the context of AI, see e.g., Nathalie A. Smuha, "Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea," *Philosophy & Technology* 24 (2020), https://doi.org/10.1007/s13347-020-00403-w.

[15] Kashmir Hill, "The Secretive Company That Might End Privacy as We Know It," *The New York Times*, January 18, 2020, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html.

[16] There is a risk that, without knowledge or consent of the persons affected, these databases may, over the longer term, may be used to track ever more aspects of people's lives, through what Evgeny Morozov calls the 'invisible barbed wire.' See Evgeny Morozov, "The Real Privacy Problem" *MIT Technology Review*, October 22, 2013, https://www.technologyreview.com/2013/10/22/112778/the-real-privacy-problem/. In Bill Davidow's words, such opaque databases may serve as a key component in the construction of an 'algorithmic prison' where the images collected may be used to profile "persons of interest, to use for facial recognition and characterisation, or even to predict a person's propensity for committing a crime" (Bill Davidow, "Welcome to Algorithmic Prison," *The Atlantic*, February 20, 2014, https://www.theatlantic.com/technology/archive/2014/02/welcome-to-algorithmic-prison/283985/).

[17] See respectively Point 4(b) and Point 7(a) of the Proposal's Annex III, to give but a few examples.

If wrongs or harms arise due to the failure of another person(s) to discharge their legal duties, the legal system must provide an effective and legitimate framework to ensure that those duties are duly enforced and that legal rights are effectively protected. This requires effective public enforcement mechanisms, including the provision of an independent judiciary. It may also require an independent public enforcement office endowed with legal powers to gather information and investigate potential violations, and to initiate enforcement action against those who appear to have failed to abide by their legal duties and obligations.

Effective judicial protection is a crucial element within the broader framework of the rule of law in the European Union. Victims of fundamental rights violations must be legally empowered to seek effective remedies, including compensation for harm against those deemed legally responsible.

Another element of the rule of law is the need for legal clarity and coherence. The various duties and obligations must be appropriately structured so that they can provide clear guidance to legal subjects concerning unlawful behaviour and its consequences. In addition, the laws themselves, when taken as a whole, must display consistency and internal coherence. The law cannot usefully serve its critical guidance role to members of the community unless its laws, taken together, are intelligible, constituting an internally coherent and consistent body of rules and principles that ground and enable peaceful coordination and co-existence among its members.[18]

A regulatory framework for Legally Trustworthy AI therefore requires an effective enforcement architecture, which establishes and protects procedural rights and is internally coherent.

c) Ensure meaningful transparency, accountability and rights of public participation *(democracy)*

Thirdly, Legal Trustworthiness requires that the regulatory framework is rooted in democratic deliberation and continuously promotes opportunities for public participation and transparency.

As discussed earlier, one of the main differences between legal and ethical standards is the extent to which institutions exist to promulgate, interpret, and enforce those standards. Moreover, in well-functioning democratic communities, significant legislative measures are preceded by open and active public debate and deliberation to establish the community's views about those measures. In democracies, the legal system endows affected stakeholders and the community at large with meaningful rights to participation in determining the legal norms which affect their collective life (both *ex ante* and *ex post*). In short, democracy requires that the regulatory framework around AI derives its legitimacy from consultation with citizens and grants them a prominent role in the design and enforcement of the Regulation.

Moreover, democracy requires that the AI systems which are allowed under the Proposal do not undermine the ideals of transparency and accountability, which are both required for meaningful public participation and democratic accountability. For example, individuals and the community at large are entitled to transparency so that they may be informed of, adequately understand, and contest the way in which AI systems make or are directly implicated in making decisions which significantly affect them.

[18] See generally Lon L. Fuller, *The Morality of Law* (New Haven: Yale University Press, 1969), 67-68; Ronald Dworkin, *Law's Empire* (Cambridge: Harvard University Press, 1986).

To summarise, Legal Trustworthiness requires (1) appropriately allocated responsibility for the harms and wrongs associated with AI, especially concerning potential *fundamental rights* implications (2) an effective enforcement framework and respect for the *rule of law* and (3) a commitment to *democracy* through meaningful rights of transparency and public participation.

4. HOW THE PROPOSAL FALLS SHORT OF THESE THREE FUNCTIONAL REQUIREMENTS

Below we demonstrate the ways in which the Proposal falls short of the three functions of Legal Trustworthiness just outlined, and thus fails to secure Legally Trustworthy AI in its current form. Firstly, we argue that the Proposal does not allocate and distribute responsibility for wrongs and harms appropriately and in a manner that adequately protects fundamental rights. We consider the Proposal's understanding of fundamental rights, its scope, prohibitions, and the provisions for biometric systems and high-risk AI (4.1). Secondly, we argue that the Proposal does not ensure an effective framework for the enforcement of legal rights and responsibilities. We discuss conformity assessments, the Proposal's coherence, the absence of procedural rights for individuals, and the Proposal's enforcement architecture and mechanisms (4.2). Thirdly, we argue that the Proposal fails to ensure meaningful transparency, accountability, and rights of public participation. We argue for public participation and information rights, highlighting substantial democratic deficit in the Proposal's standard-setting framework (4.3). Drawing on those considerations, the last chapter of this document offers concrete recommendations which we urge the Commission to consider (5).

4.1 The Proposal does not allocate and distribute responsibility for wrongs and harms appropriately, and in a manner that adequately protects fundamental rights

The following sections address the ways in which the Proposal does not adequately allocate responsibility for the wrongs and harms associated with AI – the first pillar of Legally Trustworthy AI. The way in which the Proposal operationalises fundamental rights protection appears to rest on a flawed understanding of the nature of fundamental rights. We highlight inadequacies in the way in which the Proposal understands and operationalises fundamental rights protection (4.1.1); the scope of the protection offered by the Proposal (4.1.2); the content and scope of the prohibited AI practices (4.1.3); the scope and strength of protection offered against the risks arising from the use of AI-enabled biometric systems (4.1.4); and the fundamental rights protection provided against the risks posed by 'high-risk' systems (4.1.5).

4.1.1 The Proposal operationalises fundamental rights protection in an excessively technocratic manner

Fundamental rights protection is a core aim of the Proposal (as mentioned in its explanatory texts and recitals), and one of the core elements of Legally Trustworthy AI. Yet the following sections argue that the Proposal does not provide sufficient fundamental rights protection. First, while the Proposal's text is infused with fundamental rights-language, it seems to take an overly technocratic approach to the protection of fundamental rights (a). Second, in doing so, it also overlooks the distinct nature of fundamental rights, which cannot be equated to safety standards (b). Third, the Proposal's risk categorisation of AI systems remains insufficiently stratified to safeguard individuals against the harms that they can pose (c).

a) The distinct nature of 'fundamental rights' is overlooked

The Proposal fails to properly understand the distinctive nature of fundamental rights, which requires a particular form of protection well-established in EU fundamental rights law.

To ensure due respect for the fundamental rights of all persons in virtue of their humanity, fundamental rights are accorded special weight in the architecture of legal rights protection in which rights are not merely 'interests' of individuals to be 'balanced' against the interests of others, including collective interests.[19] Treating fundamental rights as an afterthought in any 'balancing process' – whereby economic and innovation concerns take priority – means failing to recognise that fundamental rights enshrined in the Charter generate legitimate expectations for EU citizens regarding their relationship between themselves, EU institutions, Member States, and private organisations. This crucially includes *legally enforceable* safeguards against an increased scope for unchecked abuse of power and power asymmetries arising from the introduction of new technologies (between empowered users of said technologies, whether public or private, and relatively disempowered individuals). Moreover, failing to recognise the distinct nature of fundamental rights risks overlooking the Commission's obligations to respect, promote[20] and to avoid adversely affecting the rights contained within the Charter,[21] and can ultimately undermine, rather than protect, the Union values of human dignity, freedom, equality, and solidarity.

Due to their enhanced moral strength and importance, fundamental rights are accorded strong presumptive legal protection, and the justification for any fundamental rights interference therefore consists of three essential elements:

(a) First, except for rights which admit of no qualification, such as the right to freedom from torture and slavery, fundamental rights can only be interfered with in a narrow and designated range of legitimate purposes, where those interferences are prescribed by law and necessary and proportionate in a democratic society. In modern legal systems, the structure of rights protection establishes a clear framework for addressing normative conflict between rights and other important collective interests, and between rights *inter se*.

(b) Second, the duty of demonstrating this demanding burden of justification lies on those seeking to interfere with fundamental rights.

(c) Third, the determination of whether this burden of justification has been discharged is a matter for adjudication by a legitimately constituted independent body *(i.e.,* a court or body with similar powers of authoritative decision-making, such as a tribunal).

As presently drafted, the Regulation appears to treat fundamental rights as equivalent to mere interests. Each fundamental right engaged by this Proposal and the activities it enables is limited and made subject to a balancing process by the Proposal itself – at least to some degree – by virtue of (a) the inherent tension in this Proposal between the goals of promoting economic activity and innovation, with the protection of fundamental rights; and (b) the legal bases upon which it has been founded, namely, a rough amalgamation of Article 114 TFEU (which enables the Commission to propose harmonisation measures relating to the internal market) and Article 16 TFEU (which enables the introduction of harmonisation measures designed to protect the right of personal data of individuals). This creates an uneasy marriage that does not provide protection for the full gambit of potential interferences, intrusions and violations of

[19] Consider in this regard also Article 52 of the EU Charter, which provides that "Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others" (European Union, "Charter of Fundamental Rights of the European Union," Official Journal of the European Union C83 53, (2010)).

[20] European Union, "EU Charter," Article 51(1).

[21] European Union, "EU Charter," Article 54.

fundamental rights made possible by the development, deployment and use of AI systems in the EU.

Worryingly, the mechanisms through which fundamental rights are accorded protection by the proposed Regulation seem to fall seriously short of the three elements just outlined. This is particularly demonstrated by the centrality in its regulatory architecture accorded to a 'risk'-based approach, which is regarded as a largely technical and administrative matter to be attended to by AI providers by putting in place suitable 'risk management systems.' Accordingly, the Proposal's risk-based approach does not appear compatible with the regime's claim to offer a 'high level of protection' to fundamental rights due largely to a lack of systematic, meaningful scrutiny and oversight, particularly given the existing weak enforcement and remedial framework currently proposed (see section 4.2). It is taken for granted that a system of self-assessment by AI providers guarantees sufficient protection against the dangers which these technologies generate for the fundamental rights of those living in the EU.

For a legal instrument to be considered adequately rights protective – particularly where it makes significantly strong claims of the kind made in this Proposal – we would expect that it would *at least* recognise the distinctive nature of fundamental rights. This should be clearly demonstrated in the *binding* content of the Proposal itself. The Proposal needs to be aligned with existing European fundamental rights legislation and practice, which establishes substantive and procedural requirements for potential interferences with fundamental rights, including the principles of proportionality, necessity, and independent oversight.

Problems relating to fundamental rights manifest themselves in several problematic ways which are addressed throughout this document, including:

- the creation of a narrow and overly-prescriptive regime of rules for the identification of 'high-risk' systems, which is likely to produce a weak and formalist regime of rights protections;
- the creation of a system of asymmetric protection that does not offer equal protection for the fundamental rights of individuals against interferences by private organisations and public authorities;
- a lack of an independent, external body to evaluate whether rights interferences are necessary in a democratic society for pre-specified legitimate purposes and therefore proportionate;
- the provision of excessively broad discretionary power to developers and providers to determine whether their AI systems are fundamental rights compliant;
- a failure to offer affected subjects a suite of individual rights and remedies by which to obtain redress for any (actual or potential) fundamental rights interferences, as appropriate;
- weak transparency and accountability requirements which are not explicitly tied to fundamental rights protection.

In sum, the Proposal does not yet do justice to the distinct nature of fundamental rights and their need for an effective and comprehensive system of protection, as required by the first element of Legal Trustworthiness.

b) The current technocratic approach fails to give expression to the spirit and purpose of fundamental rights

Instead of taking the approach well-established in fundamental rights law, the proposed Regulation translates the protection of fundamental rights to a set of requirements, primarily

of technical nature, which should be complied with. These are insufficient to ensure that fundamental rights are effectively respected in the context of AI systems, and should hence not be seen as equivalent to the 'protection of fundamental rights.'

The Proposal's requirements erroneously reduce the careful balancing exercise between fundamental rights to a technocratic process, thus rendering the need for such balancing invisible. Currently, the Proposal claims to present:

> a balanced and proportionate horizontal regulatory approach to AI that is limited to the minimum necessary requirements to address the risks and problems linked to AI, without unduly constraining or hindering technological development or otherwise disproportionately increasing the cost of placing AI solutions on the market.[22]

Yet when considered through the lens of fundamental rights, this approach falls well short of effective rights protection – offering a weaker form of market-focused regulation as opposed to specific and clear protections against fundamental rights interference which might be generated by particular AI systems. We argue that these burdens should be reversed. Innovation is a legitimate – and perhaps necessary aim, given the remit, goals and obligations of the Commission and the legal basis upon which this Proposal has been developed. However, as argued in the previous section, internal market considerations cannot take priority over fundamental rights. Doing so significantly threatens the fundamental rights of individuals arising from the development, deployment and use of AI systems, particularly when those systems are used to inform and even to automate the exercise of decision-making power by public authorities and private companies.

A serious weakness of the regime is its failure to grapple with the highly controversial nature of AI applications, treating AI systems as analogous to other consumer products. However, while the deployment of many existing AI applications already implicates a wide range of fundamental rights, there is no systematic regulatory oversight to enable independent evaluation of whether any resulting rights interferences can be justified for the achievement of permissible purposes which are necessary and proportionate in a democratic society. The Proposal must avoid treating AI systems solely as technical consumer products which may yield claimed economic and social benefits. Instead, it needs to acknowledge that they are *socio-technical systems* which mediate social structures (and in doing so, must grapple with administrative and institutional requirements and cultures), and which can produce significant consequences for individuals, groups, and society more generally.

c) The risk-categorisation of AI systems remains unduly blunt and simplistic

Not only does the Proposal fail to take seriously the distinct nature and strength of fundamental rights, the risk-based approach taken by the Proposal also remains unduly blunt and simplistic.

Compared to the binary approach put forward in the *White Paper on AI*, the Proposal's more nuanced degree of risk strata amongst AI systems is an important improvement. A distinction is currently made between (1) prohibited practices (with certain exceptions) in Title II, (2) high-risk AI systems in Title III, (3) systems requiring increased transparency measures due to a risk of deception in Title IV (which could also fall under the high-risk systems category) and (4) all other systems. However, for several reasons, this approach may still not be sufficiently stratified in practice to adequately protect fundamental rights.

First, the Proposal does not prohibit the deployment of AI systems which violate fundamental rights other than systems which engage in the prohibited practices. Instead, AI systems

[22] European Commission, "The Proposal," Explanatory Memorandum, 3.

(including 'high-risk' systems) may be deployed even if they interfere with fundamental rights, if those deploying them adhere to the requirements set out under Title III and have in place a self-certified quality risk management system which has not been independently evaluated or reviewed.

Second, only AI systems which have specifically been identified by the European Commission as high-risk are subjected to mandatory *ex ante* requirements, such as ensuring high quality data, testing for bias, or securing human oversight. In other words, based on the Proposal, either a system is considered as 'high-risk,' or it poses no risk at all. Systems which are not specifically listed as 'high risk' might in practice pose a risk to fundamental rights, but are not subjected to these requirements, nor to an impact assessment or other mechanisms which require providers to reflect on potential risks. And while Title IV of the proposed Regulation does introduce a category of AI systems which require increased transparency measures, it only pertains to three types of systems and is limited to an obligation to 'inform' people that they are subjected to an AI system, rather than including some of the requirements of Title III. The protection afforded by the Proposal hence entirely depends on whether the closed 'high-risk list' is sufficiently comprehensive and up to date. This demand may be hard to meet, especially given the technology's evolving nature.

While the Commission's aspiration to clearly delineate high-risk AI applications is understandable from the point of view of providing legal certainty to AI providers, it fails to capture the more nuanced reality of the way AI systems are used – and have an impact on fundamental rights. Moreover, the Proposal's reliance on a 'list-based approach' also opens the door to restrictive interpretations of the AI applications that fall under its scope. AI providers who wish to circumvent the mandatory requirements imposed on high-risk AI systems may find creative ways to argue that their system does not fall under one of the narrowly defined high-risk categories, thereby pushing effectively 'high-risk' systems towards the 'no-risk' category. In other words, the lack of effective fundamental rights protection is due partly to the adoption of a highly prescriptive, list-based approach to the identification of 'high-risk' systems. Although the Proposal appears to rest on a belief that this prescriptive list-based approach will enable legal certainty, it is likely to produce the opposite effect.

Even with the possibility for the Commission to update the high-risk list by means of a delegated act, the danger remains that the list – which already merits improvement – will date rapidly and fail to provide a future-proof layer of protection against the substantial adverse impacts of the technology. To close this protection gap, an extension of the current 'list-based' approach with an approach based on broader risk-criteria, as can also be found under the General Data Protection Regulation, should hence be considered.

To summarise, the Proposal does not provide adequate fundamental rights protection, as it fails to recognise the distinct nature of those rights and the well-established doctrines on fundamental rights interferences. Its requirements for dealing with AI systems which implicate fundamental rights do not acknowledge the specific and demanding scrutiny (including the tests of necessity and proportion required by human rights law), and the risk categories specified by the Proposal may be too simplistic to provide adequate protection. As a result, responsibility for interferences with fundamental rights currently generated by AI applications (and which may arise from future AI applications) is not appropriately distributed under the Proposal, leading to a deficit in Legal Trustworthiness.

4.1.2 The Proposal's scope is ambiguous and requires clarification

The previous sections demonstrate why the Proposal's approach to fundamental rights protection is currently inadequate, resulting in a failure to appropriately allocate responsibility

for the actual and threatened wrongs and harms of AI. The next sections outline shortcomings associated with the Proposal's scope, which also affect the way in which responsibility for the harms of AI is distributed, and thus the first element of Legal Trustworthiness. We address (a) the Proposal's lack of clarity in its definition of AI, (b) lack of clarity concerning the position of academic researchers, (c) the gap in legal protection against military AI, and (d) the fact that security and intelligence agencies are not mentioned.

a) The Proposal's current definition of AI lacks clarity and may lack policy congruence

The primary objective of the Proposal is to regulate 'AI systems.' How 'AI' is defined is therefore essential for the determination of who bears legal duties under the Proposal and the scope of its protection. We argue that the definition of AI could lead to confusion and uncertainty, and although we do not propose a concrete solution in this document, we believe it merits significant attention.

Article 3(1) of the proposed Regulation states that "'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with." The techniques and approaches listed in Annex I include "(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods."

The description of software which provides outputs for human-defined objectives is incredibly broad, as it encompasses virtually all algorithms.[23] Annex I is meant to specify which particular techniques fall under 'artificial intelligence' techniques. However, the techniques mentioned seem to include virtually all computational techniques, as machine learning, inductive, deductive, and statistical approaches are all included. As the title of Annex I refers to "Artificial Intelligence Techniques and Approaches," it could be assumed that the definition of, for example, 'logic-based approaches' is limited to logic-based approaches *to artificial intelligence*. However, as 'artificial intelligence' is defined as *any* algorithm which uses the techniques listed in Annex I, this specification has become circular and is therefore not a specification at all.

For computer scientists working with any computational techniques which draw on logic or statistical insights, but which are not conventionally seen as falling within the domain of artificial intelligence, it is hard to determine whether this Regulation applies to them. This applies especially to instances of 'simple automation,' in which logic-based reasoning and probabilistic approaches are used in algorithms which simply execute pre-programmed rules and do not engage in optimisation and learning (think, for example, of virtual dice which display the number six with a certain probability). In safety-critical and fundamental rights-critical contexts, it is of little relevance whether a system relies on simple automation rather than on optimisation and learning – although the specific risks and dangers associated with the respective categories are distinct, and risk assessment procedures for the different categories of systems might need to be designed with the distinct properties of the category in mind.

[23] If we take the definition of an algorithm to be "A finite series of well-defined, computer-implementable instructions to solve a specific set of computable problems," from "The Definitive Glossary of Higher Mathematical Jargon: Algorithm," Math Vault, accessed 21 June 2021, https://mathvault.ca/math-glossary/#algo.

However, it remains the case that the providers of 'simple automation' systems will not consider their systems as being covered by the list in Annex III, as it remains a list of 'high-risk *AI* systems.' The fact that 'AI' is not clearly defined may therefore lead to legal uncertainty for providers of simple systems which use logic-based or statistical approaches which do not rely on optimisation and learning, and would therefore not immediately be thought of as 'artificial intelligence.'

There is an open question as to whether this legal uncertainty could be solved by changing the name and scope of the proposed Regulation, which could be done in two ways.

The first option consists of broadening the scope of the Proposal, and changing the name from 'Artificial Intelligence Act' to 'Algorithms Act' or 'Software Act.' This is a reasonable approach if the goal of the proposed Regulation is to guarantee the safety of products in the Union internal market and the respect for fundamental rights and Union law. From the point of view of risks to safety or fundamental rights, it is irrelevant whether a system which is "used by law enforcement authorities to detect deep fakes" (Annex III, 6(c)) is based on deep learning or on cryptography-based authentication, for example. While one system could be appropriately described as 'AI,' and the other cannot, in both cases, deepfake detection for law enforcement remains a high-risk domain of software engineering with potentially profound impacts on fundamental rights. If the risk categories are defined by reference to the social domain in which a system is used, rather than the specific computational technique used, it would make sense to extend the scope of the application to non-AI software applications in those specific domains. This would also contribute to legal certainty, as it would completely circumvent the thorny question of whether a particular software system can be said to fall under the contested notion of 'AI,' and instead focus on the domain of application, which can be specified in Annex III.

The second option for solving the issue of the legal uncertainty created by the definition of AI is changing the name and scope of the proposed Regulation in a limiting manner. Instead of broadening the scope to include all algorithms used for the purposes specified in Annex III, the Commission could limit the scope to only include systems which rely on machine learning methods. This would have advantages regarding legal certainty, as machine learning systems are much easier to define than 'AI' systems in general. However, this approach would have the crucial disadvantage of excluding from its scope many computational techniques in safety-critical and fundamental rights-critical domains, and can therefore not be recommended from the point of view of fundamental rights protection.

The current approach taken by the proposed Regulation may be attempting to find a middle ground, in that it gives the impression that it only covers a precisely defined subset of computational techniques, while in reality covering a poorly defined range of computational techniques in a more precisely defined range of social domains. This could create legal uncertainty for software providers who provide systems which operate in the domains listed in Annex III but do not necessarily think of themselves as 'AI' providers. As the determination of who is an 'AI provider' and therefore subject to the Proposal is crucial for the allocation of potential legal responsibilities to this person, the uncertainty around the definition of AI merits attention.

b) Clarification is needed on the position of academic research

The second issue relating to the scope of the Proposal and the allocation of responsibility is the position of academic researchers working within universities ('academic research'). It is currently unclear to what extent the proposed Regulation applies to academic research. The overall tone of the Proposal, which focuses on the promotion of a "well-functioning internal

market"[24] of AI in the Union, as well as the legal basis of the Proposal, seem to suggest that academic research falls outside of the scope of the proposed Regulation, to the extent that it is done in a non-commercial capacity (thereby outside the context of the *market*) and subject to its own scientific ethics standards.

However, the text of the Proposal is not clear on the position of academic researchers. Article 2(1)(a) states that the Regulation applies to "providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country." A 'provider' is defined in Article 3(2)[25] as "a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge." Article 3(11) defines "putting into service" as "the supply of an AI system for first use directly to the user or for own use on the Union market for its intended purpose."

While an academic researcher may not be a 'provider' who places an AI system on the market in the course of a commercial activity (Article 3(10)), they might be accurately described as "a natural person" who "develops an AI system with a view to" supplying it "for first use directly to the user or for own use" "free of charge." The text of the Proposal seems to suggest that in such a case, the proposed Regulation would apply to academic researchers. However, recital 16 seems to contradict this interpretation. It makes an exception to the prohibition on AI systems which deploy subliminal techniques: "research for legitimate purposes" in this area may still occur if "such research does not amount to use of the AI system in human-machine relations that exposes natural persons to harm and such research is carried out in accordance with recognised ethical standards for scientific research."

On the one hand, recital 16 seems to suggest that research generally does fall within the scope of the proposed Regulation, as it outlines specific circumstances in which the prohibition in Article 5(1)(a) does not apply (namely if no natural persons are exposed to harm and ethical principles are followed). On the other hand, this exception is not repeated in Article 5 or any of the other articles. The applicability of the Proposal to academic research must therefore be clarified.

In this clarification, it would be useful to make a clear distinction between research which takes place in academic institutions, which have well-established ethics review systems and accountability measures in case of breach of these procedures, and R&D departments in private corporations which often lack such institutionalised ethics procedures. The "recognised ethical standards for scientific research" mentioned in recital 16 can differ wildly between academia and corporate R&D. A situation in which the level of fundamental rights protection differs between these contexts would be unacceptable.

Finally, attention should be drawn to the fact that there are currently no mandatory rules for so-called 'in-the-wild' testing or experimentation with AI systems by public or private actors. If research falls outside of the scope of the Regulation, this creates a potential loophole which makes it possible for providers and users of systems which would otherwise be subject to the requirements for high-risk or prohibited systems to claim that they are merely doing 'in-the-wild' experiments, rather than actually developing or deploying a high-risk or prohibited AI system. Yet, in-the-wild testing of some AI applications – such as the trials, testing and training of autonomous vehicles on public roads, for example – is safety-critical and often occurs

[24] See European Commission, "The Proposal," 1.

[25] The Proposal says 3(1) but it comes after 3(1) and before 3(3), so it is reasonable to assume that the document meant to say 3(2).

without the consent of the affected public. Accordingly, it is vital that such activities are subject to appropriate legal regulation. This is, however, currently lacking and appears to be excluded from the scope of the Proposal.

c) The potential gap in legal protection relating to military AI should be addressed

The third issue concerning the scope of the Proposal is the potential gap in legal protection against fundamental rights interferences through military AI.

Recital 12 explains that "AI systems exclusively developed or used for military purposes should be excluded from the scope of this proposed Regulation where that use falls under the exclusive remit of the Common Foreign and Security Policy regulated under Title V of the Treaty on European Union (TEU)." Additionally, Article 2(3) states that "[t]his Regulation shall not apply to AI systems developed or used exclusively for military purposes."

The condition "where that use falls under the exclusive remit of the Common Foreign and Security Policy regulated under Title V" of the TEU is not repeated in Article 2(3). It is therefore unclear whether this condition actually applies, or whether all AI systems developed or used exclusively for military purposes are excluded from the scope of the proposed Regulation, regardless of whether they fall within the exclusive remit of the Common Foreign and Security Policy.

This difference is crucial, considering the existence of the European Defence Fund (EDF), which invests heavily in military AI.[26] The legal basis of the EDF is not Title V TEU (mentioned in recital 12 of the Proposal), but "Article 173(3), Article 182(4), Article 183 and the second paragraph of Article 188, of the Treaty on the Functioning of the European Union."[27] Article 173 TFEU falls under Title XVII, "Industry," and Articles 182, 183 and 188 fall under Title XIX, "Research and Technological Development and Space." These legal bases differ significantly from Title V TEU "General Provisions on the Union's External Action and Specific Provisions on the Common Foreign and Security Policy." This choice of legal basis for the EDF is also reflected in the fact that the European Commission plays a dominant role in the allocation of funds under the EDF, while the European Defence Agency (established under the Common Foreign and Security Policy (CFSP), Title V TEU) merely has an observer role, assisting the Commission.[28] AI projects developed in the context of the EDF therefore do not fall within the exclusive remit of the CFSP and would therefore not be excluded from the scope of the proposed Regulation if one follows the text in recital 12. However, the fact that the condition of the exclusive remit of the CFSP is not repeated in Article 2(3) of the Proposal suggests that the AI systems developed in the context of the EDF would in fact be excluded from the scope of the Regulation.

This discrepancy between recital 12 and Article 2(3) must be clarified, as the text in recital 12 seems to create an apparently arbitrary distinction between military AI systems developed in

---

[26] See Christoph Marischka, "Artificial Intelligence in European Defence: Autonomous Armament?" *The Left in the European Parliament*, January 14, 2021, https://left.eu/issues/publications/artificial-intelligence-in-european-defence-autonomous-armament/.

[27] European Parliament, "European Parliament legislative resolution of 18 April 2019 on the proposal for a regulation of the European Parliament and of the Council establishing the European Defence Fund (COM(2018)0476 – C8-0268/2018 – 2018/0254(COD))," 18 April 2019, https://www.europarl.europa.eu/doceo/document/TA-8-2019-0430_EN.html.

[28] European Parliament, "Legislative resolution of 18 April 2019," Article 28: "1. The Commission shall be assisted by a committee within the meaning of Regulation (EU) No 182/2011. The European Defence Agency shall be invited as an observer to provide its views and expertise. (…)"

the context of the CFSP and military AI systems which were developed in the context of the EDF.

Moreover, Title V TEU concerns the Common Foreign and Security Policy, which does not cover the full range of Member State foreign and security policy activities, only those foreign and security policy activities which relate to the common Union defence policy. Recital 12 therefore seems to suggest that military AI systems which were developed in the domestic context of Member State militaries, and which are used according to individual Member State policies, rather than common Union policies, would in fact be subject to the proposed Regulation. This, again, is something that the phrasing of Article 2(3) seems to contradict. Again, the condition of the exclusive remit of the CFSP in recital 12 creates confusion as to the scope of Proposal and must therefore be clarified.

In addition to the legal uncertainty created by recital 12, there are substantive concerns about the exclusion of military AI from the scope of the proposed Regulation. The HLEG provided four "examples of critical concerns raised by AI:" (1) "Identifying and tracking individuals with AI;" (2) "Covert AI systems;" (3) "AI enabled citizen scoring;" (4) "Lethal autonomous weapons systems."[29] While the Regulation deals with the first three concerns at least to an extent (in Article 5(1)(c); Article 5(1)(d); and Article 52), Article 2(3) makes it so that lethal autonomous weapons systems are excluded from the scope of the Proposal. This is concerning as the development of autonomous lethal weapons "raises fundamental ethical concerns, such as the fact that it could lead to an uncontrollable arms race on a historically unprecedented level, and create military contexts in which human control is almost entirely relinquished and the risks of malfunction are not addressed."[30]

The exclusion of military AI from the scope of the Proposal is not only worrying given the risks associated with autonomous lethal weapons, it is also problematic considering that the European Union is itself actively involved in the development of military AI systems through the European Defence Fund. This involvement includes the development of the so-called "Eurodrone" which has the capacity to be armed.[31] The Union itself could therefore be considered as actively contributing to the risks associated with autonomous lethal weapons.

The Regulation of the European Defence Fund establishes an ethics review for funding applications.[32] However, this ethics review only consists of the Commission reviewing an ethics self-assessment performed by the applicants, which is woefully insufficient considering the high-stakes nature of military AI. Moreover, the Regulation of the EDF only applies to those military AI projects which seek funding from the EDF. Military AI systems developed with other sources of funding are therefore not even subject to the minimal requirement of ethics self-assessment under Union law.

The current legal landscape means that one of the most critical concerns raised by AI identified by the HLEG is not addressed by the proposed Regulation, and is insufficiently addressed by the Regulation of the European Defence Fund. This is an unacceptable gap in legal protection, which is even more egregious considering the active Union involvement in the creation of high risks through the funding of projects like the Eurodrone.

---

29 HLEG, "Ethics Guidelines," 33.

30 HLEG, "Ethics Guidelines," 33-34.

31 Marischka, "Artificial Intelligence in European Defence," 11.

32 See Regulation (EU) 2021/697 of the European Parliament and of the Council of 29 April 2021 establishing the European Defence Fund and repealing Regulation (EU) 2018/1092, OJ L 170, 12.5.2021, 149–177, Article 7(2).

d) Clarity is needed on the applicability of the Proposal to national security and intelligence agencies

The final concern about the scope of the Proposal and how it leads to a particular distribution of legal responsibility concerns national security and intelligence agencies.

The scope of the proposed Regulation explicitly excludes AI systems exclusively developed or used for military purposes (Article 2(3)), and it explicitly includes AI systems used by law enforcement (Annex III(6)), administrative agencies dealing with migration (Annex III(7)), and judicial authorities (Annex III(8)). However, while both the excluded and the included domains can be informed by the work of intelligence agencies, the Proposal and the Annexes do not explicitly mention national security or intelligence agencies.

Article 3(40) defines "law enforcement authority" as "any public authority competent for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security." This definition could include intelligence agencies, which could be seen as a public authority competent for the safeguarding against threats to public security. This, however, depends on how 'safeguarding' is interpreted, as intelligence agencies often do not have executive powers and mainly function as a source of information for other authorities such as defence ministries, public prosecutors, immigration agencies, etc. (although this may depend on the Member State legal system). Does the collection of information amount to 'safeguarding?' Moreover, as intelligence agencies often lack executive powers, they are commonly thought of as separate from law enforcement agencies, with a "jurisdictional firewall"[33] separating them. If intelligence agencies fall under "law enforcement authorities" as defined in Article 3(40), they should be explicitly mentioned.

If intelligence agencies do not fall under 'law enforcement,' it is unclear whether AI systems developed and used for national intelligence purposes fall within the scope of the proposed Regulation, as national intelligence activities could be classified as related to military action, law enforcement, border control, and the administration of justice – depending on the specific context.

Moreover, national security is a fundamental rights-critical domain even if it does not immediately inform the high-risk domains listed in Annex III. Suppose an intelligence agency deploys a natural language processing (NLP) model which automatically screens social media posts for potential extremist content. This model flags extremist content posted by an individual. This individual is then put on a watchlist, which is shared with law enforcement agencies. The individual is now put under extra surveillance by the intelligence agency and might be apprehended by the police for having posted particular content. This NLP system would not fall under the individual risk assessment, polygraphs, deep fake detection, evaluation of the reliability of evidence, predicting the occurrence of a crime, profiling, or crime analytics listed in Annex III under 'law enforcement.' Yet, such a system affects the fundamental rights of the person subjected to it. Moreover, if intelligence agencies are not covered by Article 3(40), this might also prevent the application of the prohibition on real-time remote biometric identification (Article 5(2)) to the activities of intelligence agencies. This would cause an unacceptable gap in fundamental rights protection, as one of the goals of Article 5(2) is to

[33] Jonathan M. Fredmant, "Intelligence Agencies, Law Enforcement, and the Prosecution Team," *Yale Law & Policy Review* 16, no. 2 (1998), 331.

prevent "a feeling of constant surveillance" which dissuades "the exercise of the freedom of assembly and other fundamental rights" (recital 18).

To sum up, the scope of the Proposal significantly affects the way in which responsibility for any harms and wrongs caused by AI systems is allocated and distributed. Several aspects of the scope therefore merit clarification or improvement: (a) the definition of AI, (b) the position of academic research by university researchers, (c) military AI, and (d) national security and intelligence agencies.

4.1.3 The range of prohibited systems and the scope and content of the prohibitions need to be strengthened, and their scope rendered amendable

So far, we have argued that the Proposal reflects an inadequate understanding of fundamental rights and requires substantial clarifications regarding its scope. In the following sections, we address the list of prohibited practices, which is an important element of the Proposal's protection of rights and its allocation of responsibilities.

We welcome the inclusion of a set of 'prohibited practices' in the proposed AI Regulation. It provides much-needed legal protection against a set of AI applications and practices which are so rights-intrusive that they cannot be justified and therefore should be legally prohibited. This approach is in line with the HLEG's Policy Recommendations which supported the introduction of "'precautionary measures' when scientific evidence about an environmental, human health hazard or other serious societal threat (such as threats to the democratic process), and the stakes are high."[34] By offering concrete legal protection to individuals and communities against the wrongs and harms which arise from such practices, the prohibitions have the potential to significantly strengthen the legal protection of the fundamental rights they implicate.

However, we argue that the way in which these prohibitions are currently drafted is deficient. In particular, for various reasons set out below, we argue that the scope of the list of prohibited practices should be revised to (a) strengthen the prohibition on manipulative practices, and (b) clarify the provisions on AI-enabled social scoring. Moreover, the Proposal remains too tolerant of rights-intrusive biometric applications which also require stronger prohibitions. The subject of biometrics is however dealt with later (under section 4.1.4), since these systems fall partly under prohibited practices and partly under high-risk applications.

a) The scope of prohibited AI practices should be open to future review and revision

The Proposal currently lists prohibited AI practices in Article 5. In contrast to the list of stand-alone high-risk systems in Annex III, the list of prohibited systems in Article 5 of the Proposal cannot be amended by the European Commission. The current list of prohibited practices seems heavily inspired by recent controversies, especially the prohibitions on social scoring and remote biometric identification by law enforcement authorities. The problematic nature of certain uses of AI can sometimes only be grasped when those uses are actually put in practice (e.g., much of the controversy around social scoring systems emerged in the context of recent, concrete developments in China). However, the fact that some practices have received recent media attention does not mean that they are the only AI practices which are deeply problematic. Future uses of AI systems can be hard to predict, and it seems premature to permanently fix the list of prohibited AI practices.

[34] HLEG, "Policy and Investment Recommendations," 38, section 26(2).

Against this background, we recommend considering an option for the European Commission to add prohibited practices to Article 5 following review and consultation, in particular with the European Parliament. This review mechanism, which should ensure due respect for legal certainty, could be triggered by reference to a set of criteria (set out in Title II) which would allow the Commission to revise and update the scope of prohibited practices. A requirement for wide public consultation prior to the addition of prohibited practices should be considered, to ensure that public debate on the matter is duly reflected in the Regulation.

These additions would also strengthen the current weaknesses of overall fundamental rights protection in the Proposal explained earlier, and contribute to the appropriate allocation of responsibilities in the context of AI. Indeed, if it is impossible to add AI practices to the list of prohibited practices over time, it might not be possible to allocate appropriate responsibility for the wrongs and harm ensuing from potentially damaging future AI practices. Recourse to the lengthy procedure of the adoption of a new regulation by the European Parliament and the Council (which could take several years) risks leaving individuals unprotected from future AI practices which could pose an unacceptable level of risk to fundamental rights.

b) Stronger protection is needed against AI-enabled manipulation

The prohibition on subliminal manipulation is also under-protective, as it only applies to the exploitation of a limited set of vulnerabilities, leaving the door open to many non-subliminal manipulative AI practices. Moreover, it does not impose any obligations for uses of subliminal techniques which could be considered tolerable if undertaken with the full and informed consent of the affected target in a carefully monitored and rigorously supervised manner.

First, we question the choice regarding the practices outlined in Article 5(1)(a) and (b) to concentrate on only physical and psychological harms. Other significant, but excluded harms include financial and economic harms, cultural harms, harms of recognition and autonomy harms, but also collective and societal harms.[35] There is no reason to hold that the harms following from the use of manipulative AI should be limited only to the first two types, or that they are necessarily more serious. Manipulative technologies which interfere with a person's fundamental rights and lead to significant harm of all sorts ought to be prohibited. We therefore recommend that references to physical and psychological harm are removed to be replaced simply by 'harm and fundamental rights interference.'

Second, we are concerned by the concentration of attention in the practices outlined in Article 5(1)(a) and (b) on harms as they might be experienced by an individual person. Some of the harms and wrongs outlined above may extend beyond individuals to groups of people and society as a whole. Certain applications of AI are best judged as impacting 'society;' for instance, algorithmically driven processes used on social media sites might have negatively impacted democratic engagement.[36] In cases like these, we can observe harm which is difficult to discern and demonstrate on an individual level, but still has far reaching adverse societal implications. Accordingly, in addition to the alterations suggested in the previous paragraph, the Commission could consider extending the prohibition's provision to include 'harm to groups and to collective interests and values,' with 'values' referring to the Union values indicated in Article 2 of the Treaty of The European Union (TEU).

---

35 See also Victoria Canning and Steve Tombs, *From Social Harm to Zemiology* (London: Routledge, 2021), Chapter 3.

36 See, for instance Siva Vaidhyanathan, *Anti-Social Media: How Facebook Disconnects Us and Undermines Democracy* (Oxford: OUP, 2018); Nathaniel Persily and Joshua Tucker (Eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform* (Cambridge: CUP, 2021), various chapters.

Third, Article 5(b) of the Proposal prohibits the use of AI systems "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." This Article highlights the vulnerabilities of persons on account of age, physical or mental disability. However, the logic behind the choice of only these protected characteristics – and the exclusion of characteristics such as race, sex, religion and ethnicity – which are all protected characteristics under EU equality law, is deeply puzzling and apparently unjustified. Age and mental disability could be read as indicating a question of mental capacity, but the inclusion of physical disability confounds this. Rather than limit the characteristics this way, we recommend that the clause should be expanded to include all grounds listed in Article 21 of the EU Charter on Fundamental Rights.

Fourth, there are manipulative AI practices which do not rely on subliminal techniques. Article 5(1)(a) only prohibits 'subliminal' manipulation, or AI systems which "deplo[y] 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." Subliminal techniques are methods of "presenting information below the subjective threshold of awareness."[37] While it is debated exactly how much these methods can act as a stimulus for action, there is general agreement that, to some degree, they do. This definition is echoed in recital 16 which mentions "subliminal components individuals cannot perceive," which indicates that subliminal techniques cannot be perceived. The same is reflected in the Proposal's Explanatory Memorandum, which states that the "prohibitions covers (sic) practices that have a significant potential to manipulate persons through subliminal techniques beyond their consciousness."[38] However, apart from the fact that the prohibited practices are not perceivable and that they are 'beyond consciousness,' it is not entirely clear what 'subliminal' means in the context of AI systems.

'Subliminal' is usually understood as referring to sensory stimuli which are too subtle to consciously perceive. Yet, AI systems can be manipulative and cause harm without relying on sensory stimuli which are impossible to perceive. Consider a chatbot which learns ways of tricking people into revealing their passwords, for example. This chatbot does not emit sensory stimuli which are impossible to perceive – its statements can easily be seen, read, and consciously processed by humans. However, the result of those statements might still be that someone is manipulated into sharing their passwords. Another example is social media disinformation which plays on people's anxieties by misrepresenting facts with a view to influencing their behaviour is manipulative.[39] Ostensibly the individual can act as they please, but once a person's deeper concerns are triggered, it becomes less clear what that means. This is manipulation which may 'materially distort a person's behaviour in a harmful manner' but which may not be subliminal at all.

It is unclear whether such manipulative systems which do not rely on subliminal cues would be covered by the prohibition in Article 5(1)(a). This could lead to legal uncertainty. Moreover, it would leave a gap in responsibility for manipulative AI practices which do not rely on subliminal cues. Consequently, we suggest that where AI systems employ technologies that

---

37 Mary Still and Jeremiah Still, "Subliminal Techniques: Considerations and Recommendations for Analyzing Feasibility," *International Journal of Human–Computer Interaction* 35, no. 5 (2018), 457-466. See also Sid Kouider and Stanislas Dehaene "Levels of Processing during Non-Conscious Perception: A Critical Review of Visual Masking," *Philosophical Transactions: Biological Sciences* 362, no. 1481 (2007), 857-875.

38 European Commission, "The Proposal," Explanatory Memorandum, 12.

39 Caroline Jack, "Lexicon of Lies: Terms for Problematic Information," *Data & Society Research Institute*, August 9, 2017, https://datasociety.net/library/lexicon-of-lies/.

manipulate individuals (including, but not limited to, systems that utilise subliminal techniques) in a way that causes harm or interferes with their fundamental rights should be prohibited.

Finally, there may be uses of subliminal technologies which influence people's actions, but which are not illegitimate or harmful. There are examples of subliminal manipulation which aid motivation and could be self-consciously chosen by the user with their full and informed consent.[40] However, for this to be so, there needs to be a power relationship between the 'manipulator' and 'manipulated' which is close to symmetrical, such that the 'manipulation' is plainly transparent. With this in view, we recommend that subliminal techniques to which those affected have given their free and informed consent and which do not cause harm or fundamental rights interference are excluded from the prohibition, although they must be subject to the transparency and disclosure obligations stated in Title IV on Transparency Obligations for Certain AI Systems, Article 52.

c) The provisions on AI-enabled social scoring need to be clarified and potentially extended to private actors

In addition to the problems regarding the fixed nature of the list of prohibitions and the seemingly arbitrary particulars of the prohibition on manipulation, the prohibition on social scoring also needs to be clarified and extended.

Article 5(1)(c) of the Proposal prohibits AI systems to be used 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 (…) detrimental or unfavourable treatment of certain natural persons or whole groups (…)" and/or "treatment (…) that is unjustified or disproportionate to their social behaviour or its gravity." In short, it prohibits AI-enabled social scoring by public authorities.

However, social scoring is often conducted by private actors, who potentially have access to enormous amounts of personal data, covering many important areas of life, like job applications, hiring policies, and loan applications. Social scoring models in these domains could have devastating effects. Furthermore, private organisations are increasingly moving into areas of social policy which would have previously been occupied by the state, and employ social scoring models to identify areas of need, like the identification of children most at risk of being mistreated with a view to taking them into care against the wishes of their parents, or the children themselves. In this context, it is crucial to clarify what 'on behalf of' public authorities' means, as some social domains occupied by the state in one Member State could be in private hands in another. This could potentially undermine the Proposal's goal of a 'coordinated European approach' to the risks associated with AI.

Secondly, Article 5(1)(c) focuses on "social behaviour or known or predicted personal or personality characteristics," while it is well documented that proxies may be employed for where such personal data is protected. These proxies could be drawn from factors not mentioned in the Article, such geographical location (postcodes, etc.). Scoring on these measures can be as discriminatory and devastating for individuals as those drawn from the characteristics included in the Proposal. This should be reflected in the language of the Article.

[40] Henk Aarts, Ruud Custers, and Martijn Veltkamp, "Motivating Consumer Behaviour by Subliminal Conditioning in the Absence of Basic Needs: Striking Even While the Iron is Cold," *Journal of Consumer Psychology* 21, no. 1 (2011), 49-56.

In conclusion, we warmly welcome the list of prohibited practices in Article 5, as it is a crucial part of the prevention of harms and the allocation of responsibility under the Proposal, and therefore contributes to Legally Trustworthy AI. However, we recommend the inclusion of a procedure to enable review, consultation and revision to account for future uses of AI which are incompatible with fundamental rights. Moreover, the provisions regarding subliminal manipulation and social scoring must be clarified and strengthened.

4.1.4 The adverse impact of biometric systems needs to be better addressed

AI-enabled biometric systems are one of the primary regulatory targets of the Proposal. While the moratorium on facial recognition technologies mentioned in a leaked version of the *AI White Paper* has not materialised,[41] we welcome the EU's determination to tighten the control over such intrusive technologies, even redlining some uses of biometric systems. However, as we will argue in this section, the Proposal does not seem to take the fundamental rights implications of biometric technologies sufficiently seriously.

This is especially problematic because biometric systems can pose threats to several fundamental rights enshrined in the EU Charter. These affected rights are not limited to the rights to privacy and data protection, which are straightforwardly implicated by the use of biometric systems. Less obvious implications of biometric systems include interferences with the right to non-discrimination which may be contravened by biased systems, or systems which categorise people on the basis of physical features which are implicitly or explicitly related to protected characteristics such as age, sex, ethnicity or race. Moreover, widespread implementation of biometric systems in society also risks generating chilling effects with significant implications for the rights to free speech and free assembly and therefore ultimately the foundations of democracy itself.[42] In addition, the 'invisible' or 'silent'[43] nature of these systems and the difficulty of avoiding particular public spaces make it so that genuine transparency is hard to achieve.

As discussed in section 4.1.1, fundamental rights have an enhanced moral and legal status, and cannot simply be overridden for the sake of convenience or efficiency. Indeed, fundamental rights are accorded strong presumptive legal protection, such that any interference with those rights must be justified by those seeking to interfere with them in accordance with heightened standards of scrutiny. Those wishing to employ biometric systems ought to justify the use of such systems with reference to their legitimate purpose, necessity, and proportionality. The burden of justification is on the providers and users of biometric systems – it is not on those who are subjected to them to prove that their use was unjustified. Moreover, the final judgment on the adequacy of such a justification must lie with an independent judicial authority.

While the Proposal seems to apply this logic to a very limited category of biometric systems, it presumes that most uses of biometric systems are merely 'high-risk,' meaning that the AI provider does not need to engage in any justificatory discourse, as long as the requirements for high-risk systems are complied with. This results in a biometric-tolerant regime, which leaves

---

41 Samuel Stolton, "LEAK: Commission Considers Facial Recognition Ban in AI 'White Paper.'" *EURACTIV.Com*, January 17, 2020, https://www.euractiv.com/section/digital/news/leak-commission-considers-facial-recognition-ban-in-ai-white-paper/.

42 The Guardian Editorial, "The Guardian View on Facial Recognition: A Danger to Democracy," *The Guardian*, June 9, 2019, https://www.theguardian.com/commentisfree/2019/jun/09/the-guardian-view-on-facial-recognition-a-danger-to-democracy.

43 Lucas Introna and David Wood, "Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems," *Surveillance and Society* 2, (2004), 177, 181-2.

excessive discretion to AI providers regarding most biometric systems and fails to offer meaningful and effective protection of fundamental rights.

In what follows, we first explain the distinctions made by the Proposal between the various kinds of biometric technologies (a). Then, we critique the approach taken by the Proposal regarding biometric systems, including the limited nature of the prohibition on certain uses of biometrics (b), the unjustified exclusion of biometric categorisation and emotion recognition from the prohibition and the high-risk category (c), and the distinction made between public and private use of biometric systems (d).

a) Different types of biometric systems under the Proposal: an overview

Before we critique how the Proposal regulates biometric technologies, it is useful to outline the different categories of biometric systems envisioned in the Proposal. First, the adjective 'remote' is used to indicate only those biometrics that collect data in a passive, remote manner while excluding traditional ones requiring physical contact, e.g., fingerprints and DNA samples. Second, biometrics used for identification or verification (known as 'remote biometric identification systems,' or RBIS) are differentiated from systems which classify or categorise people (known as 'biometric categorisation systems,' or BCS). Additionally, emotion recognition systems (ERS) are separately defined and regulated despite their reliance on biometric data. Such a differentiation renders RBIS subject to regulation of a higher threshold whereas BCS and ERS are only subject to transparency obligations and not necessarily regarded as high-risk.

A distinction is also made between 'post' (retrospective) and 'real-time' (live) use of biometric systems, defined in Article 3. Both are classified as high-risk and subjected to requirements such as logging capabilities and human oversight.[44] Further, the uniqueness of their high risks renders them subjected to third party conformity assessment rather than to self-assessment, as is the case for other high-risk AI systems.[45] The only normative difference is that real-time biometrics are considered 'particularly intrusive'[46] and prohibited if used in publicly accessible spaces for the purpose of law enforcement, subject to several exceptions.[47]

b) The 'prohibition' on biometrics is not a real 'prohibition'

The use of RBIS in public spaces for the purpose of law enforcement is a prohibited practice under Article 5(1)(d). However, the prohibition is so narrow and allows for such broad exceptions that it barely deserves to be called a 'prohibition.'

Firstly, the Proposal does not mention the use of live biometrics in public spaces by public actors for non-law enforcement purposes, such as intelligence gathering for the purposes of national security or resource allocation, or for migration management purposes. The prohibition therefore only covers a limited range of uses of biometric systems, and by a very limited range of actors (i.e., law enforcement). Regardless of whether it is used by law enforcement or by other public (or private) actors, the use of this technology implicates fundamental rights in a way that is almost inevitably disproportionate, as it requires the processing of a great amount of sensitive data of many people to enable the identification of few individuals. The prohibition should therefore be extended beyond the use of RBIS by law

44 European Commission, "The Proposal," Recital 33.
45 European Commission, "The Proposal," 14.
46 European Commission, "The Proposal," Recital 18.
47 European Commission, "The Proposal," Recital 19 and Article 5(1)(d).

enforcement, and *at least* also cover its use by public authorities (or private actors acting on their behalf) that have any coercive power over individuals.

Secondly, broad exceptions exist to the already limited prohibition on the use of RBIS. Article 5(1)(d) allows law enforcement to use RBIS in public spaces if such use is strictly necessary for one of the following objectives: (1) the targeted search for specific potential victims of crime, including missing children; (2) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack; or (3) the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA and punishable in the Member State concerned by a custodial sentence of a detention order for a maximum period of at least three years, as determined by the law of that Member State. These exceptions are allowed as long as the "seriousness, probability and scale of the harm caused in the absence of the use of the system" and the "consequences of the use of these systems for the rights and freedoms of all persons concerned" are taken into account.

Although we recognise that there might in principle be circumstances which render the use of RBIS by law enforcement in public spaces justifiable in a manner that conforms with respect for fundamental rights, the scope of the current exceptions does not meet the requisite standard. In particular, the third ground of exception, which allows for the use of such technology for the detection of perpetrators or suspects of criminal offences covers a vast number of situations. This places excessive discretion in the hands of law enforcement agencies. Moreover, the consequences of permitting RBIS in public spaces means that infrastructures to enable the permitted use of this technology will be built and rolled out across Members States at scale. Once these infrastructures are built, function creep and potential misuses or abuses of such an infrastructure remain a very real concern. It is important to stress that the risks relating to this infrastructure go beyond an impact on individual privacy, but risks affecting collective values and the integrity of democracy at large. Merely knowing that the infrastructure exists, and not being able to ascertain whether it is currently being used, may lead to severe chilling effects on the exercise of political rights such as the freedom of speech and freedom of association.

As the 'prohibition' on biometric systems only applies to a very specific type of biometric system; a limited set of purposes for the system; and a limited set of actors, and as very broad exceptions to the prohibition exist, still enabling the construction of a biometric surveillance infrastructure, we believe that this 'prohibition' barely deserves its title, and fails to provide adequate fundamental rights protection.

c) The Proposal does not take a principled approach to the risks of various biometric systems

The prohibition on the use of RBIS by law enforcement in public spaces just discussed, is justifiable from a fundamental rights perspective. Yet, it is unclear why the prohibition in Article 5(1)(d) does not include BCS or ERS.

Firstly, the range of possible legitimate uses of remote live BCS for law enforcement purposes in public places is very difficult to imagine. Differently than in the case of RBIS, this technology does not aim to search for a specifically identified – potentially at danger or dangerous – individual, but for a group of people. Indeed, the premise behind BSC is that people can be categorised based on their physical characteristics, such as visible gender, race, age, and the way they move their bodies. As these physical characteristics relate closely to the protected characteristics outlined in outlined in Article 21 of the Charter of the Fundamental Rights of the EU, any use of such technology for law enforcement purposes is suspect, and requires justification based on its necessity and proportionality in a democratic society, provided for by law. Moreover, these technologies present serious risks to minorities, as one

could easily imagine a system which is designed to recognise members of religious minorities, racial minorities, or members of the LGBT+ community, which could be seen as 'useful' information for law enforcement purposes.[48] It is not clear why the Proposal presumes that RBIS is always more intrusive or damaging than BCS. Therefore, we recommend that the use of remote live BCS in public spaces for law enforcement purposes (and other public authorities with coercive powers) should be prohibited outright and admit of no exceptions.

As regards the use of BCS *ex post*, given its intrusive and inherently discriminatory nature, it would be important to ensure that any permissible use thereof is subject to strict safeguards – including *ex ante* control and *ex post* review by an independent authority. Therefore, all other uses of BCS which are not captured by the afore-suggested prohibition should be included in the list of high-risk AI systems, and subjected to *ex ante* conformity assessment and additional safeguards in terms of necessity and proportionality, as well as transparency, rule of law and due process.

Secondly, it is unclear why ERS are seen as less intrusive than RBIS. The premise behind ERS is that emotions can be read from human bodies, which can then be used for interventions. Yet the idea that emotions can be read from human bodies in any scientific and objective way has long been debunked: it is based on 'pseudoscience' rather than established scientific evidence.[49] It is impossible to understand how a pseudoscientific technique could ever be properly regarded as necessary and proportionate for legitimate law enforcement purposes. Moreover, even if ERS were scientifically sound, automatically reading every single person's emotional state is a highly intrusive act, with potentially huge chilling effects, which again seems unjustifiable for law enforcement purposes. We therefore urge the Commission to prohibit any use of ERS for law enforcement purposes (and other public authorities with coercive powers), and admit of no exceptions given the lack of any scientific soundness of this technology and its extremely intrusive nature.

As regards the use of ERS for other purposes – for instance, deployed by private actors – we believe they should at the very least be included in the list of high-risk AI systems. Moreover, the Commission should consider subjecting ERS intended to be used in one of the high-risk domains to *ex ante* conformity assessment and control, in particular when used in situations with power asymmetries or on vulnerable individuals and groups.

d) The distinction between private and public uses of remote biometric systems requires justification

Not only does the Proposal not seem to adopt a principled approach to the regulation of the different types of biometric systems, it also does not justify the distinctions it makes between different users of biometric systems.

Currently, individuals are offered an asymmetrical level of protection against the use of AI-enabled biometric recognition systems. While the use of RBIS in public spaces by law enforcement authorities is prohibited, its use by private organisations is merely categorised as 'high-risk.' The Proposal thus creates a 'two-tier' system of regulation by which state

---

[48] Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81 (2018), 1–15. David Leslie, "Understanding Bias in Facial Recognition Technologies: An Explainer," *Alan Turing Institute*, October 26, 2020, https://doi.org/10.2139/ssrn.3705658.

[49] Evan Selinger, "A.I. Can't Detect Our Emotions," *OneZero*, April 6, 2021, https://onezero.medium.com/a-i-cant-detect-our-emotions-3c1f6fce2539. Luke Stark and Jesse Hoey, "The Ethics of Emotion in AI Systems," October 9, 2020, https://osf.io/preprints/9ad4u/. Douglas Heaven, "Why Faces Don't Always Tell the Truth about Feelings" *Nature*, February 26, 2020, https://www.nature.com/articles/d41586-020-00507-5.

authorities are held to a higher standard, to the detriment of rights protections for citizens from private actors. Recital 18 justifies this on the basis that:

> The use of AI systems for 'real-time' remote biometric identification of natural persons in publicly accessible spaces for the purpose of law enforcement is considered particularly intrusive in the rights and freedoms of the concerned persons, to the extent that it may affect the private life of a large part of the population, evoke a feeling of constant surveillance and indirectly dissuade the exercise of the freedom of assembly and other fundamental rights.

However, these same problems may exist with private use of biometric technologies. It is not immediately clear why this approach differs from that of the General Data Protection Regulation, wherein no distinctions are made between public and private 'data controllers,' recognising that AI systems expand and extend the powers of private companies and opening fundamental rights to new forms of abuse.[50]

This is not to say that the Proposal could not legitimately distinguish between the use of biometrics by public authorities and by private actors. Nonetheless, we would expect at least a justification as to why such a distinction is necessary, and why government use of biometric systems poses a greater risk to fundamental rights than use by private organisations. Otherwise, these restrictions seem to operate on the assumption that the use of biometrics systems by private organisations are somehow less intrusive, despite their potential for widespread use in 'semi-public' spaces such airports, shopping centres, and sports stadiums. Personal data collected by private actors could be shared with law enforcement bodies, opening up a worrying loophole in the protections afforded by the proposed Regulation.

The Proposal's approach seems to overlook the fact that private biometric systems are likely to be widespread and may also cause a 'feeling of constant surveillance;' and that private actors may share information with law enforcement authorities or collaborate with them. It is true that the justification for the interference with fundamental rights traditionally comes from public actors. However, considering the enormous power of private actors in the space of AI and their ability to directly contribute to chilling effects on the exercise of fundamental rights, the Commission should consider tightening the rules for private use of biometric systems.

In sum, to ensure that the development and use of AI-enabled biometric technologies in the EU are Legally Trustworthy, the Proposal needs to take seriously the duty to justify any interference with fundamental rights according to the principles of fundamental rights law. To enable such a discourse of justification, the Proposal requires a more principled approach towards biometric technologies, which includes strengthening the current prohibition of RBIS, prohibiting live remote BCS in public places as well as the use of ERS by law enforcement and other public authorities with coercive powers, and reconsidering the existence of different regimes for public and private actors. Moreover, uses of ERS and BCS which are not included in the suggested prohibition should be added to the list of high-risk systems of Annex III, and subjected to strong safeguards. Only then can the Proposal move towards more adequate prevention of the harms associated with biometrics, and the appropriate allocation of responsibility for such potential harms.

4.1.5 The requirements for high-risk AI systems need to be strengthened and should not be entirely left to self-assessment

After having considered the proposed regulatory frameworks for prohibited systems and biometric systems, we now turn to high-risk AI systems as defined by the Proposal.

---

[50] Linnet Taylor, "Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector," *Philosophy and Technology* (2021), https://doi.org/10.1007/s13347-020-00441-4.

Title III of the proposed Regulation sets out a new regulatory regime with mandatory requirements for AI systems that pose a high risk to the health and safety or fundamental rights of natural persons – namely high-risk AI systems.[51] Rather than defining the concept of 'high risk,' the Proposal specifically lists which systems fall under this category, which we have referred to as a prescriptive 'list-based approach.' High-risk systems include systems intended to be used as safety components of products which are subject to third party ex-ante conformity assessment covered by EU legislation listed in Annex II of the Proposal, and other stand-alone AI systems used in high-risk domains listed in Annex III. The former concern AI systems which are, for instance, covered by the Machinery Directive (2006/42/EC), the Toy Safety Directive (2009/48/EC) or the Medical Devices Regulation (2017/745). The latter concern a limited list of AI applications in various areas, including in the field of education, HR, public services and law enforcement.

The mandatory requirements for high-risk AI systems are broadly inspired by the 'requirements for Trustworthy AI' listed in the HLEG's Ethics Guidelines and must be complied with prior to the system's placing on the market or putting into service. They pertain more particularly to data quality and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy, and security. The imposition of such mandatory requirements is a considerable step forward in advancing the protection against the adverse effects of AI systems. At the same time, the Proposal should still be substantially revised regarding the way in which high-risk systems are defined via a prescriptive list-based approach, and the requirements which apply to high-risk systems.

a) Outsourcing the 'acceptability' of 'residual risks' to high-risk AI providers is hardly acceptable

As argued in section 4.1.1, the Proposal takes a rather technocratic approach to fundamental rights, imposing a list of obligations on the providers of high-risk AI systems, rather than making them engage with the justificatory discourse customary in human rights law. Not only does this choice poorly reflect the spirit of fundamental rights, it also confers undue discretion for the AI provider.

For high-risk AI systems, Article 9 of the Proposal mandates the establishment, implementation and documentation of a risk management system to be maintained as an iterative process throughout the AI system's lifecycle. The risk management system should *inter alia* include the identification and analysis of known and foreseeable risks, an estimation and evaluation of risks which may emerge when the system is used in accordance with its intended purpose and under conditions of 'reasonably foreseeable misuse,' and the adoption of 'suitable' risk management measures. These measures, according to Article 9(3), "shall give due consideration to the effects and possible interactions resulting from the combined application of the requirements" and "take into account the generally acknowledged state of the art, including as reflected in relevant harmonised standards or common specifications." In short, the effectiveness of the mandatory requirements imposed on high-risk AI systems thus hinges on the quality of this risk management system and the protection which it is intended to provide.

Importantly, the Proposal goes beyond a mere requirement of risk management documentation, and also requires that systems are tested so as to identify the necessary risk management measures to ensure compliance with the various requirements (Article 9(5)). In addition, while the risk management process primarily falls upon the provider of the AI system, the provider

[51] European Commission, "The Proposal," Explanatory Memorandum, 13.

needs to anticipate the knowledge, experience, education and training that can be expected from the user, as well as the environment in which the system is intended to be used.

At the same time, the Proposal appears to leave an unduly large amount of discretion to the provider of the AI system as regards the execution of the risk management process.

Article 9(4) leaves it up to the AI provider to determine which measures to take in order to ensure that "any residual risk associated with each hazard as well as the overall residual risk of the high-risk AI systems is judged acceptable." This means that the decision about which risks are deemed 'acceptable' is outsourced to the AI provider, who also seeks to put the system on the market or into service. Bearing in mind that this includes not only risks to the health and safety of individuals, but also risks to fundamental rights, this outsourcing appears seriously problematic. The fact that the AI provider needs to communicate those residual risks to the user does not offer much solace. At the very least, we wonder why there is no obligation for the AI provider to consult with stakeholders, such as those who will be subjected to the AI system or otherwise have a legitimate interest, about which level of risk may be deemed 'acceptable.'

Reference can be made to Article 35(9) of the GDPR which, in the context of a data protection impact assessment, states that "where appropriate, the controller shall seek the views of data subjects or their representatives on the intended processing, without prejudice to the protection of commercial or public interests or the security of processing operations." By analogy, a similar provision to ensure the involvement of those affected by the AI application when establishing the risk management system and setting out mitigation measures should hence also be considered here.

In the same vein, the fact that testing procedures shall be "suitable to achieve the intended purpose of the AI system" and need not go beyond what is necessary to achieve that purpose, and that testing thresholds should be set which are "appropriate to the intended purpose of the high-risk AI system," leaves it up to the AI provider to decide how the words 'suitable' and 'appropriate' are interpreted. Given that the system is only (potentially) subjected to independent control *ex post*, this margin of discretion is difficult to justify, especially as it may pertain to – sometimes difficult – judgment calls in a fundamental rights-sensitive area. This is even more so for high-risk AI systems which are already on shaky grounds due to a lack of any scientific evidence for their appropriateness, such as emotion recognition systems, or systems which involve highly vulnerable individuals such as refugees and children. In sum, the discretion left for AI providers and deployers is reflective of the Proposal's deficient approach to fundamental rights outlined above.

b) It can be questioned why the listed high-risk AI systems are considered acceptable at all

Besides the fact that providers of high-risk AI systems have too much discretion to decide on the acceptability of those systems, one could wonder why these systems – with clear and potentially severe adverse implications for fundamental rights – are considered to be acceptable at all. Their categorisation as posing a 'high' yet nevertheless 'acceptable' risk (since if unacceptable, they would figure under the prohibited AI practices of Article 5) is especially blatant given that no evidence is provided that their adverse impact on fundamental rights is necessary and proportionate in a democratic society, as required by human rights law (see section 4.1.1).

In this regard, reference can also be made to the joint opinion of the European Data Protection Board and the European Data Protection Supervisor concerning the Proposal. In this opinion[52], it is stressed that biometric categorisation technology in public spaces – whether used by public authorities or private entities – based on ethnicity, gender, political or sexual orientation, or other discrimination grounds prohibited under the Charter seems incompatible with the fundamental right to data protection and should hence be categorised as prohibited rather than high-risk. The same point is made regarding "AI systems whose scientific validity is not proven or which are in direct conflict with essential values of the EU,"[53] such as AI-enabled polygraphs, which are currently included under Annex III and hence deemed to pose a high yet acceptable risk to fundamental rights according to the Commission.

Furthermore, and as already mentioned in section 4.1.4(c), it is also unclear why emotion recognition systems, which according to the Study supporting the Impact Assessment of the AI regulation "lack scientific reliability and validity"[54], so that "any sentiment analysis software attempting to recognise human emotions is thus unproven,"[55] are not even considered as posing a 'high risk' despite the calls for a ban on such technology, unless when used by public authorities in migration cases or by law enforcement. It is, in our view, at least doubtful that this categorisation is reflective of the informed public opinion on this matter, and that the impact of such systems on fundamental rights can be legitimised as necessary and proportionate.

To appropriately allocate responsibilities for the risks of high-risk systems, the list of high-risk applications cannot include AI practices which are incompatible with fundamental rights and whose use cannot be reasonably justified, such as technologies which are manifestly discriminatory and/or lacking in any clearly established scientific basis.

c) The adaptability of the Scope of Title III is too limited

As noted above, the scope of Title III – i.e., high-risk AI systems – is limited to those systems which are covered by Annex II of the Proposal (intended to be used as a safety component of a product, or a product covered by the specifically listed Union harmonisation legislation) or by the list of stand-alone high-risk systems laid down in Annex III. This Annex provides a list of eight areas in which, according to the European Commission, high-risk AI applications are used. Within each of these areas, the Annex lists a limited number of AI applications which are considered as high-risk and hence subjected to the mandatory requirements laid down in Title III.

Pursuant to Article 7(1) of the Proposal, the European Commission can amend the list in Annex III and include additional systems if it can demonstrate that those systems pose a risk of harm to the health, safety or fundamental rights of individuals which is "in respect of its severity and probability of occurrence, equivalent to or greater than the risk of harm or of adverse impact posed by the high-risk AI systems already referred to in Annex III." Furthermore, Article 84(1) mentions that the Commission will review this list on a yearly basis, in order to keep it up to

---

52 European Data Protection Board, "EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)," 18 June 2021, https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf.

53 European Data Protection Board, "Joint Opinion," 12.

54 Andrea Renda et al., "Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe," April 21, 2021, *Publications Office of the European Union,* https://op.europa.eu/en/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1/language-en/format-PDF/source-204305195.

55 Renda, "Study," 39.

date with technological developments. Crucially, however, the Commission is only empowered to add high-risk AI systems to Annex III if they "are intended to be used in any of the areas listed in points 1 to 8 of Annex III." This means that the list of high-risk AI systems is only amendable to the extent that new high-risk applications fall under the already existing headings mentioned in Annex III.

This is problematic for two reasons. Firstly, there are already categories of AI systems which should be classified as high-risk, but are not currently listed in Annex III, nor would they fall under any of the eight headings. For example, high-frequency trading algorithms have profound impacts on the market and have the potential to have destabilising effects on economies.[56] These algorithms do not fall under any of the listed areas and can therefore never be classified as high-risk systems. Secondly, there might be categories of high-risk systems which we are currently not aware of, considering that the field of AI moves fast and increasingly permeates other fields. For example, whereas the field of cyber security was not conventionally seen as an 'AI field,' it is now flooded with AI research, and many cyber security solutions now rely on AI. There could be similar developments in other fields, which we might not currently be able to foresee. For these two reasons, it would be advisable to foresee the opportunity to include new domain categories to the list in Annex III.

Finally, while recital 85 of the Proposal seems to indicate that the Commission should also have the power to amend Annex II (namely the list of Union legislation covering systems that pose a risk to safety and hence fall under the high-risk category) – no such powers seem to be granted to the Commission in the Proposal's articles. Instead, the Commission's delegated powers to amend the list of high-risk AI systems seem to be limited to an update of Annex III. This omission might be an oversight on behalf of the drafters of the Proposal and would need to be corrected prior its adoption.

d) The list of high-risk AI systems for law enforcement should be broadened

Even when leaving aside concerns about the future-proofness of the list-based approach to high-risk AI systems and about the potential miscategorisation of such systems and lack of policy-congruence,[57] questions already arise regarding the comprehensiveness of Annex III, specifically in the domain of law enforcement.

Annex III(6) lists the high-risk systems in the domain of law enforcement. The list is mostly focused on systems which have natural persons as their subjects ("individual risk assessments for natural persons;" "detect the emotional state of a natural person;" "profiling of natural persons"). By focusing on natural persons, the Annex fails to identify optimisation systems which use geospatial data to determine law enforcement resource deployment and/or prioritisation as high-risk systems (otherwise known as 'predictive policing,' or 'crime hotspot analytics systems'). Generally, such technologies are used to produce statistical predictions regarding where future crime may take place, in order to enable law enforcement authorities to 'optimise' where and how they deploy their resources for maximum benefit. Resource optimisation systems do not solely rely on personal data of natural persons, but draw from a wide range of data relating to a geographical location, including the occurrence and rate of occurrence of crimes in a specific geographical location.

---

[56] See generally Maureen O'Hara, "High-Frequency Trading and Its Impact on Markets," *Financial Analysts Journal* 70, no. 3 (2018).

[57] In this regard, see also Karen Yeung's submission to the public consultation on the European Commission's White Paper on Artificial Intelligence of February 2020, which elaborates on the problems relating to a list-based approach to high-risk AI systems. The submission can be accessed here: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3626915.

Despite not relying on personal data of natural persons, the fundamental rights implications of these systems are important because they are used to determine who can be subject to increased police intervention (based on geographical location), how these interventions occur, and with what frequency. Used for law enforcement purposes, without due care, resource optimisation systems may contribute to over-policing and surveillance of specific geographical locations caused by data 'feedback loops,'[58] and in doing so, may further exacerbate existing problems with systemic discrimination arising from historical racial and socio-economic biases in existing policing datasets (as geographical location is often a proxy for race or economic class), with little opportunity for redress, and no transparency for affected individuals. Drawing from Philip Alston speaking about the SyRI welfare fraud detection system in the Netherlands, targeting entire neighbourhoods as suspect and "subject to special scrutiny" with a combination of digital and physical methods threatens the very essence of privacy, by contributing to general unease, potential prejudice, and chilling effects on behaviour.[59] It matters little whether the personal data of natural persons is implicated here.

Were the final text of the Regulation therefore to exclude resource optimisation systems, this could amount to a significant failure to recognise the systemic social assumptions that are built into AI systems – as socio-technical systems which mediate social institutions and structures – which by virtue of their implementation, further mediate the enjoyment of individuals' fundamental rights, including respect for their human dignity, equality, liberty and other freedoms. We therefore recommend that geospatial AI systems are included under Annex III(6), and that reference is made to AI systems which affect the distribution of law enforcement resources.

e) The requirements that high-risk AI systems must comply with need to be strengthened and clarified

Our final remarks on the proposed regulatory framework for high-risk AI systems concerns the strength and clarity of the requirements for such systems. While vague language – which might cause legal uncertainty and a weak protection against AI's adverse effects – can be found under several requirements for high-risk systems, we highlight a few questions and concerns regarding three requirements in particular: those pertaining to data governance, transparency and human oversight.

Data governance obligations

The first requirement which raises questions is 'data governance.' Firstly, the large discretion for providers of high-risk AI systems mentioned earlier is also reflected in the requirements for data governance. Article 10 of the Proposal, dealing with requirements of data quality and governance, also lets the term 'appropriate' do some heavy lifting. Indeed, it requires that training, validation and testing data sets shall be subject to 'appropriate' data governance and management practices, that the data sets shall have 'appropriate' statistical properties, and that the processing of special categories of data to avoid the risk of bias is carried out subject to 'appropriate' safeguards for fundamental rights. While the Article can be commended for specifying the minimal considerations that should be taken into account for the data management process to be considered 'appropriate,' it leaves open what constitutes an

[58] The Law Society of England and Wales, "Algorithms in the Criminal Justice System," June 4, 2019, https://www.lawsociety.org.uk/en/topics/research/algorithm-use-in-the-criminal-justice-system-report, 35.

[59] Philip Alston, "Brief by the United Nations Special Rapporteur on extreme poverty and human rights as *Amicus Curiae* in the case of NJCM c.s./De Staat der Nederlanden (SyRI) before the District Court of the Hague (case number: C/9/550982/HA ZA 18/388)," *OHCHR*, September 26, 2019, https://www.ohchr.org/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf, 29.

appropriate statistical property. Does this require that the data be a representative sample of the entire population, or merely of the potential ad hoc groups that may be subjected to the AI system's analysis or scoring? This decision is left to the AI provider. Going the other extreme, the Article also seems to require that training, validation and testing data sets shall be "complete" and "free of errors" – a potentially unrealistic obligation.

Secondly, the data governance obligations themselves do not seem to prevent the use of bad proxies. Article 10(2) requires that the assumptions about the information that the data supposedly measures and represents – which we understand to include the manner in which the data is used as a *proxy* for something else – be rendered explicit, and that potential data gaps or shortcomings are identified at the outset. Rendering assumptions and proxies explicit is an important step forward. At the same time, this does not yet in and of itself prevent the use of misguided proxies, especially given that these are not part of the information that needs to be made public in the EU database for stand-alone high-risk AI systems according to Article 60. Nothing in the Proposal seems to, for instance, prevent public authorities from using *arrest data* as a proxy for *crimes committed* (while not all arrested persons are charged or convicted, and many crimes occur for which no arrests are made). Given that these assumptions are not publicly accessible, their misguided nature may not easily come to light. It is hence difficult to argue that this provision can ensure the adequate protection of fundamental rights. Accordingly, this provision should be strengthened with a requirement to ensure that the assumptions made are reasonable and adequate, and that they are part of the information that is communicated both to the user of the high-risk AI system as well as those who may be affected thereby.

Finally, there are open questions about the new data governance requirements. For instance, interaction with the GDPR - and more specifically the applicability of the GDPR's requirements for collecting and processing data - is not addressed in the Proposal (except in the Explanatory Memorandum which states that the Proposal is without prejudice and complements the GDPR).[60] Furthermore, while Article 10(3) states that "*training, validation and testing data sets shall be relevant, representative, free of errors and complete,*" no mention is made of data integrity (in terms of data provenance). One can hence wonder what the status is of data that has been collection in violation of people's rights outside the EU – such as data from the Chinese population with less extensive privacy rights – and whether the use thereof would still be deemed appropriate by the proposed Regulation. We would hence advise the inclusion of a specific provision in the Proposal to exclude the use of such non-integer data.

Transparency obligations

Transparency regarding AI systems is intended to be one of the main fundamental rights protections afforded by the Proposal. Besides establishing a category of AI systems which require increased transparency, regardless of their risk-level (under Title IV), and the creation of the EU database for stand-alone high-risk AI systems that will be publicly accessible (under Title VII), transparency requirements are also found in Title III on high-risk AI systems. Under this Title, Article 13 mandates that a high-risk AI system be designed in such a way as to be "sufficiently transparent to enable users to interpret the system's output and use it appropriately"[61] and that it should be accompanied by instructions for use with information about the system that is "relevant, accessible and comprehensible to users." Pursuant to Article 13(3), this information should *inter alia* include the identity and contact details of the provider of the AI system, the technical specifications and details relating to the system's performance

[60] European Commission, "The Proposal," Explanatory Memorandum, 4. See also section 4.2.2 further below, concerning the consistency of the proposed Regulation with data protection legislation.

[61] European Commission, "The Proposal," Recital 32.

(including its limitations), the human oversight measures that were put in place, as well as residual risks to health, safety or fundamental rights based on the systems intended use or reasonably foreseeable misuse. However, there are three two significant problems with the approach taken to transparency here.

Firstly, the Proposal focuses specifically on transparency for the *user* of the system rather than for the individual who is *subjected* to the system. There is hence no direct engagement with the needs of EU citizens and residents, or with those individuals coming into contact with providers and users of AI systems who are based within the EU. This point is further stressed under section 4.3.2, which deals more extensively with the Proposal's deficiency in terms of transparency rights for individuals.

Secondly, as is the case for the other requirements of Title III, the use of terms such as 'sufficiently' transparent and 'appropriate' type and degree of transparency, grants a lot of interpretative leeway to AI providers who will be self-assessing their systems' level of transparency. As noted above, while Article 13(3) does mention the minimal elements which should be disclosed on to users, this does not concern information about the way in which those who are subjected to the system may be adversely impacted by the system. Furthermore, the list of information to be disclosed seems to rely on the idea that results generated in the lab, for instance in terms of performance and accuracy, will suffice. It should however be kept in mind that these results may differ substantially when it comes to 'in the field' settings that can be more unstable or different from the lab. It would hence be important to include a statement about the conditions 'in the field' in which the AI system is intended to be used, as well as about the parameters of performance testing. The lack of standardised protocols for such testing – and for quality and performance metrics – is problematic, since these are essential to ensure that the imposed transparency obligations remain meaningful outside of the lab.

Human oversight obligations

In addition to the requirements above, the requirement of human oversight also needs further clarification. Article 14 expresses an admirable desire to ensure that there is a human failsafe 'in the loop' to prevent harm or rights violations. Yet in many instances, a meaningful failsafe is impossible to secure in practice, given that the entire premise of data mining is aimed at generating insight that is beyond the capacity for human cognition. This inevitably also means that the human being who needs to exercise oversight over the system will often not be able to second-guess the validity of the system's outputs, except in limited cases where human intuition may detect obvious failures or outliers. Moreover, the problem of automation bias is unlikely to be overcome through this provision.

Further, it remains unclear whether the human oversight measures set out in Article 14 apply to the *user*, or someone *independent of the user*, or indeed whether the user refers to the organisation that uses the AI system on the whole, or to a specific individual who is responsible for a specific decision.[62] Oversight is required for all action related to the development, deployment and use of AI systems, to ensure that fundamental rights are protected to the highest standard possible. This includes human oversight of the design process *and* regular independent human oversight of those humans-in-the-loop who are responsible for taking the final decision, informed by the output of an AI system. It is not enough to know that these individuals are aware of the potential for automation bias, it must be *transparently demonstrated and ensured* that decisions are not made through over-reliance on the output of an AI system. Therefore, we recommend that under Article 14(3), a *third* category is introduced which recognises the need for *users* to implement *organisational non-technical* measures to

[62] European Commission, "The Proposal," Recital 38.

ensure robust human oversight, which consists of *at least:* training for decision-makers, logging requirements, and clear processes for *ex post* review and redress.

Finally, a clarification is needed regarding Article 14(5), which imposes an additional oversight requirement when biometric identification systems are used. In such case, the system's user cannot take any action or decision based on the identification resulting from the system, unless the result was verified and confirmed by at least two natural persons. While such strengthened oversight sounds laudable, two points can be raised. First, for this provision to be meaningful, the confirmation given by the 'two natural persons' should be based on a separate assessment (with one 'on the ground,' for instance, to sight the individual in question) rather than being reduced to two people looking at the same computer screen. Second, reliance on human oversight as a safeguard should only be used as a last resort once the use of such intrusive systems has been proven to be necessary and proportionate in a democratic society, and not as a legitimation of the use of technologies that should in fact not be used in light of their rights-violating nature. Human oversight is not a panacea for the problems that certain AI systems might introduce, and should hence not be used as an excuse for their deployment where there is no basis to do so.

The danger of over-reliance on the outputs of an AI system is best evidenced through the *Viogén* system, used in Spain to predict and prevent the risk of domestic violence against women. Fourteen out of the fifteen women who were killed in a domestic violence incident in 2014 had previously reported their aggressor, yet had been classified by Viogén as being at low or non-specific risk.[63] This shows that depending on the context of the system, even decisions resulting in 'low risk' classification can produce significant dangers, where the decision-maker does not have the requisite technical knowledge to robustly understand how the system works and to consider its limitations.

In light of the above, we therefore suggest the Commission to strengthen the protection afforded by the mandatory requirements for high-risk AI systems, and to clarify the open questions they raise.

The previous sections considered the ways in which the Proposal could be amended to come closer to attaining the first element of Legally Trustworthy AI – the appropriate allocation of responsibilities for harms caused by AI, particularly regarding fundamental rights. We addressed the Proposal's conception of fundamental rights, its scope, the content of the prohibitions, the regulatory framework around biometrics, and high-risk systems. The following sections address the second pillar of Legally Trustworthy AI: an effective enforcement framework which promotes the rule of law.

4.2 The Proposal does not ensure an effective framework for the enforcement of legal rights and responsibilities (rule of law)

As explained in Chapter 3, one of the functions of law is to allocate and distribute responsibility for harms and wrongs in society. The previous sections explained that the Proposal does not allocate responsibilities in ways which adequately protect against fundamental rights infringements. Additionally, Chapter 3 argued that the distinctive character of *legal* (as opposed to *ethical*) rules lies in an effective and legitimate framework through which legal rights and duties are enforced. The following sections therefore comment on the enforcement framework proposed in the proposed Regulation.

[63] AlgorithmWatch, Automating Society Report 2020, *AlgorithmWatch*, 2020, https://automatingsociety.algorithmwatch.org/, 227.

We argue that the current oversight, monitoring and enforcement regime falls woefully short of the standard that Legal Trustworthiness requires. As a result, the Proposal is in danger of providing the façade of legal protection, while in practice offering little meaningful and effective protection and collapsing into little more than self-regulation.

Firstly, the enforcement architecture relies heavily on (self-) conformity assessments. This leaves too much discretion to AI providers in assessing risks to fundamental rights without meaningful independent oversight, leaving many safety-critical and fundamental-rights critical AI applications without any *ex ante* review or systematic *ex post* review (4.2.1). Secondly, insufficient attention has been paid to the coherence and consistency of the Proposal, both internally and in relation to other legal instruments (4.2.2). Thirdly, the Proposal is completely silent on the enforcement of individual rights. No procedural rights are granted to individuals affected by AI systems (who are indeed rarely mentioned) and no complaints mechanism is foreseen (4.2.3). Finally, we argue that the enforcement mechanism relies too much on national competencies and is overly complex (4.2.4).

4.2.1 The Proposal unduly relies on (self-) conformity assessments

It is well-recognised that the accordance of excessive discretion to governmental officials is a serious threat to the rule of law.[64] This also applies when excessive discretion is delegated to those who are legal subjects in combination with insufficient guidance as to how to use such discretion – leading to a potential deficit in Legal Trustworthiness. As already argued in mentioned in section 4(1)(5)(a), the Proposal nevertheless provides a high level of discretion not only to public officials using high-risk AI systems, but also to private actors – as it depends almost exclusively on self-assessments to ensure that fundamental rights are complied with.

Below, we raise concerns relating to the enforcement mechanism envisioned in the Proposal with regard to the overly broad discretion which the Proposal grants to AI providers (a), and the nature of the current CE marking regime, which lacks an *ex ante* control mechanism for fundamental right-sensitive applications (b).

a) The Proposal leaves an unduly broad margin of discretion for AI providers and lacks efficient control mechanisms

Section 4.1.1 argued that the Proposal understands fundamental rights protection in an overly technocratic manner and grants too much discretion to AI providers regarding the balancing exercise involved in fundamental rights protection. These concerns relate to the way in which the Proposal allocates responsibility, but they also relate to the way in which the Proposal is enforced.

As argued section 4.1.5.a, the Proposal gives AI providers and users very broad discretion to determine what they consider to be respect for fundamental rights, given that there is no requirement for (*ex ante*) independent verification and certification that the system is in fact fundamental rights compliant. Although independent authorities are empowered to review the training, validation and test data and associated technical documentation (per Article 64), they only have residual power to inspect and test the technical and organisational systems if the documentation is insufficient to ascertain whether a breach of fundamental rights has occurred. It is difficult to understand the basis upon which an authority could establish whether the data offered for inspection and associated documentation is in fact the basis upon which the AI

[64] Per A.V. Dicey: "No man is punishable or can be lawfully made to suffer in body or goods, except for a distinct breach of law established in the ordinary legal manner before the ordinary courts of the land," in A.V. Dicey, *Introduction to the Study of the Law of the Constitution* (New York: Liberty, 1982), 23.

system is configured and whether, in practice, the system operates in a manner that respects fundamental rights (other than in the case of manifest and egregious violations) in the *absence* of technical testing.

Furthermore, many of the protections offered by the Proposal rely heavily on provider self-assessment. This includes, for instance, the 'conformity assessment' practices of Articles 19 and 43, and the 'quality management systems' of Article 17. As previously explained, the acceptability of adverse implications of AI systems on fundamental rights depends on proportionality and necessity tests through which a limited range of legitimate justifications can be considered. Yet, it remains unclear who determines what is necessary and proportionate, whether this occurs during the certification and conformity assessment processes, and whether there are means to dispute the proportionality assessment of a given system. The Proposal offers little guidance on how proportionality is considered in the development, deployment and use of AI systems, other than the responsibilities of 'notified bodies' in relation to the certification procedure (Article 44(3)). While much of the enforcement of the Proposal relies on self-assessment by AI providers, little information is given as to their responsibilities regarding fundamental rights proportionality and necessity tests.

This lack of guidance on proportionality and necessity tests gives rise to significant ambiguity, for example in relation to the definition of AI systems as high-risk, which requires consideration of its intended purposes. The designation of risk for a specific technology involves consideration of (a) whether a system *may* cause harm, and (b) whether this is *likely*. The intended purpose is defined by the AI provider and reasonably foreseeable misuse refers to the activity of the user. Article 9 provides for a 'risk management system' intended to be a "continuous iterative process run throughout the entire lifecycle of a high-risk AI system," to minimise the risk posed to fundamental rights by a given AI system. However, several questions remain open regarding its application in practice.

For example, how does this approach prevent against unintentional misuse? Are there cases in which a user may legitimately 'misuse' a particular AI system for to *protect* fundamental rights (could the user then change the intended purpose of an AI system without incurring the obligations of a provider under Article 28)? If so, who decides these thresholds? In cases where interferences with fundamental rights are in question, how are normative trade-offs negotiated to prevent any interference with, and conflicts between, different fundamental rights?

Consider, for example, a 'recidivism risk' assessment system created for the purpose of identifying teenagers who require specific and heightened therapeutic intervention to reduce their future risk of 're-offending.' Yet in practice, the results of this risk assessment significantly impact their access to public assistance benefits, and/or vocational and educational institutions – even though they have never been convicted of a crime – thus generating a double (or triple) risk with regard to those high-risk uses identified in Annex III.[33] Would this constitute misuse, if the provider identified the intended purpose as *only* being for use in recidivism risk assessment? Or would the authorities in this case be proportionately pursuing a legitimate aim? How is this tension and ambiguity resolved, given that the result significantly impacts the rights of the individual involved, including rights of the child, their right to privacy and the presumption of innocence? Even a single AI system can raise serious questions regarding a wide range of potential scenarios, risks and adverse effects on fundamental rights. It is far from clear to us that the average AI provider has the expertise to consider all the complex ways in which their systems could affect fundamental rights.

Considering the questionable ability and legitimacy of AI providers to engage in these complex balancing exercises, the self-assessment process risks amounting to a technical 'tick box' exercise. In other words, it is not a guarantee of a high standard of fundamental rights

protection. It is a system through which the relative risks might be considered and registered, which heavily depends on the actions of individual actors, who may be incentivised to accord heavier weight to economic concerns than to fundamental rights protections. More clarity is required as to how this risk would be averted.

Finally, AI providers are responsible for regulation through 'post-market monitoring.' It is unclear how this would work in a law enforcement or national security context, where there are significant obstacles concerning confidentiality and the handling of complex and sensitive data. There are therefore very serious questions as to whether monitoring would be possible, given that providers simply may not be able to obtain access to the necessary information, and therefore, further processes that ensure independent oversight of sensitive AI systems may be necessary.

In addition to AI providers, also users of high-risk systems are granted unduly wide discretion under the Proposal. Users are responsible for a similar self-assessment of risks under Article 29, which contains very little in the way of obligations beyond following the 'instructions of use' given by the provider and to maintain automatically generated logs for "a period that is appropriate in the light of the intended purpose of the high-risk AI system and applicable legal obligations under Union or national law."[65] This says nothing of the problem of 'automation bias,' or relative liability where something goes wrong. Automation bias is briefly mentioned in Article 14(4)(b) but is not substantively addressed.

In sum, the unduly broad discretion accorded by the Proposal to AI providers and users should be addressed in order to enhance legal certainty and secure more effective protection against the adverse effects raised by the use of AI.

b) The conformity assessment regime should be strengthened with more *ex ante* independent control

In addition to granting excessive discretion for AI providers, the Proposal's enforcement mechanism puts undue faith in the effectiveness of conformity assessment and CE marking. The Proposal relies heavily on these instruments in relation to high-risk systems to provide the necessary level of assurance to individuals and the public that the system will not violate their fundamental rights or threaten their safety. Yet the CE certification system and conformity procedures involving notified bodies provides weak protection of human health and safety, judging by the experience of the Medical Devices regime which allowed fraud and corruption in the medical devices industry to occur undetected for a significant period of time, including the sale of over 400,000 PIP silicone breast implants worldwide which were manufactured using cheap industrial grade silicon in violation of the CE mark.[66] Given this experience of conformity assessment conducted by Notified Bodies, it is seriously questionable whether reliance on self-certification provides meaningful legal assurance that the requirements to obtain a CE mark in relation to 'high-risk' AI systems are properly met.

At the very least, if the conformity assessment and CE marking is retained, *ex ante* verification by an independent body should be considered for a broader set of high-risk AI systems, beyond just biometric systems. If these systems are truly considered to pose a high risk to safety and fundamental rights, the lack of *ex ante* independent auditing and evaluation is an enigma. This is even more so given the severity of the impact of some of these systems on fundamental rights, especially when used in the context of vulnerable groups or when based on unscientific

[65] European Commission, "The Proposal," Article 29(5).

[66] Sylvia Kierkegaard and Patrick Kierkegaard, "Danger to public health: medical devices, toxicity, virus and fraud," *Computer Law and Security Review* 29, no.1 (2013), 13-27.

approaches such as in the case of emotion recognition or polygraphs, as commented in section 4.1.5.a.

Exacerbating this problem is the fact that the documentation and logging requirements for high-risk AI systems risk being a mere *ex post* verification of a paper trail based on self-certification rather than an effective audit. This again underlines the need for a more fine-grained approach to the various risk-levels that AI systems pose, and heightened attention to the use of these systems by governments. The presently still overly binary categorisation of a system as either high-risk or no-risk and – except for biometric identification – as either left to a self-assessment or no assessment at all, seems untenable. It should hence be explored whether certain high-risk systems listed under Annex III would benefit from a conformity assessment carried out by an independent entity prior to their deployment to ensure that their 'trustworthy' status hinges on more than a paper document.

4.2.2 There is currently insufficient attention to the coherency and consistency of the scope and content of the rights, duties and obligations that the framework seeks to establish

The delegation of significant parts of the enforcement of the proposed Regulation to AI providers and users is not the only concern which affects the rule of law – the second pillar of Legal Trustworthiness. The lack of effective and meaningful protection accorded to fundamental rights resulting from the Proposal's current formulation of a 'risk-based' approach also fails to ensure that EU laws are internally consistent and coherent because the level of protection offered falls below the standards set out in the EU Charter of Rights and Freedoms. Although we warmly welcome the attempts made in the proposed Regulation to ensure that there is clarity about how AI systems which already fall within existing EU laws will be treated, there are several legal instruments and legal doctrines which are implicated by the development and deployment of AI systems, the relationship of which to the proposed AI Regulation is not always clearly identified.

After outlining concerns regarding the Proposal's internal coherency (a), we briefly discuss the Proposal's relationship with the GDPR (b), the Law Enforcement Directive (c) and the MiFID II Regulation (d).

a) The consistency of the Proposal with EU (fundamental rights) law should be ensured

To ensure that the legal rights and obligations arising under the proposed Regulation are coherent and consistent with existing laws, the locus of responsibility for the protection of fundamental rights must be clarified. The Proposal states that one of its primary aims is to ensure a high level of protection for fundamental rights across the EU. However, the Charter obligations are primarily binding on Member States and EU institutions while the proposed Regulation imposes various legal requirements on all those deploying or putting AI systems into service. There is a lack of clarity concerning the obligation to ensure that those deploying 'high-risk' systems have quality management systems in place to ensure respect for fundamental rights extends to mechanisms that guard against rights interferences arising from the activities and actions other than those of Member States and EU authorities. In this respect, we recommend that the Proposal be amended along the lines of the GDPR, which explicitly confers legal obligations pertaining to the collection and processing of personal data on all 'data controllers' irrespective of whether the data controller is a public or private person (see also section 4.1.4.d).

Unless the legal obligations to demonstrate respect for fundamental rights mean ensuring that all actors, whether state or non-state, are required to demonstrate respect for fundamental rights by all persons involved in deploying, using and putting on the market AI systems, the draft

Regulation is unlikely to provide adequate protection of fundamental rights. We therefore recommend that the proposed Regulation be amended to make it explicit that the obligation to respect fundamental rights when AI systems are put into service applies to private sector actors in a direct and unmediated way, and not merely as legal obligations against states accorded under conventional international human rights law. The imposition of legal duties on private actors is particularly important given the serious asymmetry of power between those who are directly affected by AI systems, and the organisations with the resources and expertise to deploy them, given the capacity of these systems to operate automatically, at scale and in real time.

Additionally, it is unclear whether the proposed Regulation establishes a framework of maximum harmonisation binding in its entirety on Member States, or whether it is one of minimum harmonisation[67] that allows Member State legislatures to decide whether to set more demanding standards than those stipulated – for example, to decide to prohibit biometrics by law enforcement agencies and private use in public spaces given that there is no demonstrable and robust evidence that the use of these technologies substantially improves public safety and security.

b) The Proposal's relationship with the General Data Protection Regulation should be strengthened

The Proposal is designed to enable "full consistency with existing Union legislation applicable to sectors where high-risk AI systems are already used or likely to be used in the future."[68] In addition to fundamental rights law, this includes both the *General Data Protection Regulation* and the *Law Enforcement Directive*, both of which should complement with a new "set of harmonised rules applicable to the design, development and use of certain high-risk AI systems and restrictions on certain uses of remote biometric identification systems."[69] However, from a fundamental rights perspective, it remains to be seen whether these new rules *complement* existing standards of protection, or risk undermining them.

For instance, the GDPR contains stringent protections for 'special category data' for which collection and processing is conditional on significant thresholds of protection, while the Proposal does not offer analogous protection. Under the proposed Regulation, Article 1(b) prohibits the following:

> 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 harm.

As already mentioned in section 4.1.3.a, there is an open question about why this provision does not include *all* categories of special category data, recognised in the GDPR as requiring a higher standard of protection.

Furthermore, questions can be raised about the way in which the Proposal deals with systems relying on biometric data – which are also covered by the GDPR – as discussed under section 4.1.4 above.

At the same time, the European Commission did consider the relevance of the GDPR to this Proposal elsewhere. For instance, Article 10(5) which deals with data governance requirements

---

67 This point was also rightfully raised in Michael Veale and Frederik Zuiderveen Borgesius, "Demystifying the Draft EU Artificial Intelligence Act," *SocArXiv*, July 6, 2021, doi:10.31235/osf.io/38p5f.

68 European Commission, "The Proposal," 4.

69 European Commission, "The Proposal," 4.

for high-risk AI systems does foresee a basis to process special categories of personal data (referred to in Article 9(1) of the GDPR), to the extent this is strictly necessary to ensure the monitoring, detection and correction of bias, and subject to appropriate safeguards for the fundamental rights and freedoms of natural persons. Moreover, the Proposal provides that users of high-risk AI systems shall use the information provided to them under Article 13 to comply with their own obligation to carry out a data protection impact assessment under Article 35 of the GDPR.[70] Furthermore, in the context of the provisions on establishing regulatory sandboxes, the Commission explicitly foresees that the Proposal can provide the legal basis for the use of personal data collected to develop "certain AI systems in the public interest within the AI regulatory sandbox."[71]

Given the reliance of many AI systems on personal data, it is recommended to strengthen the ties between the Proposal and the GDPR more consistently, in order to ensure a more coherent and comprehensive data protection framework for AI systems. More generally, consistency between these two regulatory instruments covering strongly related technological practices would contribute to the rule of law, and therefore to Legally Trustworthy AI.

c) The Proposal's relationship with the Law Enforcement Directive should be clarified

In addition to uncertainties stemming from the Proposal's relationship to the GDPR, there are questions regarding its relationship to the Law Enforcement Directive[72] (LED), including (a) the duties and obligations of private organisations which develop and deploy AI systems for law enforcement use; (b) the relationship between data protection impact assessments under the LED, and conformity assessments under the Proposal; and (c) the role of safeguards provided for by the LED, which have no equivalent in the text of the Proposal.

Firstly, the duties of private AI providers in the context of law enforcement must be clarified. Private actors acting on behalf of law enforcement authorities (for the purposes of prevention, investigation, detection, or prosecution of criminal offences, or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security) are explicitly included as 'competent authorities' under Article 3(7)(b) of the LED. This is of crucial significance considering that many law enforcement authorities may procure and use AI systems developed by private organisations. We must ask then, if 'providers' (found in the Proposal) are captured by these same requirements, considering that the design of a given AI system significantly influences both (a) the procedure of law enforcement decision-making (e.g., providers of machine-learning recidivism risk assessment make important choices regarding the relevance of specific data sets, the weighting of specific features, and the form and content of decision outputs provided to law enforcement, in addition to the level of technical transparency about the logic of the system itself) and (b) the kinds of decision that can be made (prospective, future-oriented predictions based on large data sets that would not be possible by human judgement alone). Similarly, considering the obligations of providers to ensure post-market monitoring – which again, may significantly affect the form and content of decisions made – is there a threshold at which this would be considered as the 'exercise' of

[70] European Commission, "The Proposal," Article 29(6).

[71] European Commission, "The Proposal," Recital 72.

[72] European Parliament and Council, "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, 4.5.2016, p. 89–131" ("LED"), 2016.

public authority? The answers to these questions are vital for accountability for potential interferences with fundamental rights in the law enforcement context.

Secondly, the relationship between data protection impact assessment under the LED and impact assessments under the Proposal would need some clarification. Article 27 of the LED states:

> where data processing involves new technologies, and taking into account the nature, scope, context, and purposes of the processing is likely to result in a high risk to the rights and freedoms of natural persons, data controllers must produce a public Data Protection Impact Assessment which contains at least (taking into account the rights and legitimate interests of the data subjects and other persons concerned):[73]
>
> - a general description of envisaged processing operations;
> - an assessment of risks to the rights and freedoms of the data subjects;
> - the measures envisaged to address those risks;
> - safeguards;
> - security measures and mechanisms and to demonstrate compliance with this directive.

Questions remain as to how the conformity assessments referred to in the Proposal differ from data protection impact assessments, the latter of which include mention of 'safeguards,' and contribute to the ability of subjects to obtain remedies (laid out below), in the sense that data subjects have recourse to seek a judicial remedy where impact assessments (a) have not been completed; (b) are incomplete; or (c) are complete but demonstrate incompatibility with data protection requirements. This possibility of remedy has not been replicated by the conformity assessment procedure, yet not alternative has been provided.

Thirdly, the LED offers other remedies not available in the Proposal, where such assessments do not meet requirements, or obligations are not met in other ways. For example:

- Article 52 provides for a right for the data subject to lodge a complaint with the domestic supervisory authority where a breach of data protection occurs or is suspected;
- Article 53 provides for a right to an effective judicial remedy against a supervisory authority for data subjects who have lodged complaints under Article 52 and who disagree with the resultant decision or action;
- Article 54 provides for a right to an effective judicial remedy against a controller or processor for data subjects, where they believe that rights laid down in the LED have been infringed upon in a non-compliant fashion. This right is provided without prejudice to any alternative administrative or judicial remedy.

Nothing of the sort is offered by the Proposal, nor even suggested. It is assumed that such remedies can be accessed by other means, but this should *at least* be spelled out in the Proposal itself. Individuals whose fundamental rights are affected by AI systems must be provided with a clear path to access justice. As will be stressed further below, without any mention of remedies, it is unclear whether the Proposal can achieve the second pillar of Legally Trustworthy AI.

[73] European Parliament and Council, "LED", Article 27.

d) Concerns around the Proposal's implicit harmonisation with MiFID II should be addressed

Finally, there is no direct indication of the Proposal's harmonisation with the Markets in Financial Instruments Regulation (MiFID II).[74] This aspect risks representing a deficiency in the Proposal, given the growing use of algorithms in the context of multiple financial activities and services (e.g., algorithmic trading and robo-advisory) and given the fact that practices like algorithmic trading are not listed as a high-risk domain under Annex III.

The Proposal merely provides that financial authorities should be designated as competent authorities for the purpose of supervising the implementation of this Regulation and that the conformity assessment procedure should be integrated into existing procedures under the Directive on prudential supervision – namely, Directive 2013/36 / EU (with some limited derogations).[75] These indications do not seem sufficient to ensure legal certainty and consistency. Indeed, MiFID II contains several provisions that seem to overlap with the content of this Proposal. Article 17, for example, provides a notion of algorithmic trading and delegates the preparation of particular risk management practices (not based on conformity assessment and CE marking) to the European Security and Markets Authority (ESMA) to protect public savings and market stability. In requesting that the conformity assessment takes place within the framework of the Directive on prudential supervision (focusing on the role of the European Banking Authority), the Proposal might create a conflict between different risk assessment procedures and an overlap of functions between different European agencies. It would hence be important for the Commission to take this risk of inconsistency into consideration.

4.2.3 The lack of individual rights of enforcement in the Proposal undermines fundamental rights protection

In the sections above, we argued that the enforcement of the Proposal overly relies on conformity (self-) assessments and that its coherence with different parts of EU law must be ensured. In this section, we draw attention to the fact that a significant part of the enforcement mechanism is completely missing from the Proposal, namely individual rights of redress and an accompanying complaints mechanism.

One of the Proposal's aims is protecting the fundamental rights of individuals, yet these individuals do not feature in the Proposal at all. Its provisions instead focus on the obligations of the AI 'provider' and 'user,' who are often already in an asymmetrical power relation to those individuals whom they subject to their systems. This asymmetrical representation of the different stakeholders in the Proposal creates an enforcement architecture which potentially threatens the rule of law. We therefore argue that the Proposal must guarantee procedural rights to redress for individuals subjected to AI systems (a), and a complaints mechanism dealing with potential violations of the Regulation or infringements of fundamental rights (b). Later, in section 4.3.2, we also argue for more substantive rights for individuals in order to address the blatant absence of 'ordinary people' from the Proposal.

---

[74] European Parliament and Council, "Regulation (EU) No 600/2014 of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Regulation (EU) No 648/2012, OJ L 173, 12.6.2014," 2014, 84–148.

[75] See European Parliament and Council, "Directive 2013/36/EU of the European Parliament and of the Council of 26 June 2013 on access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms, amending Directive 2002/87/EC and repealing Directives 2006/48/EC and 2006/49/EC, OJ L 176, 27.6.2013," 2013, 338–436.

a) The Proposal does not provide any rights of redress for individuals

Those whose rights have been interfered with by the operation of AI systems (whether high-risk or otherwise) are not granted legal standing under this Proposal to initiate enforcement action for violation of its provisions, nor any other enforceable legal rights for seeking mandatory orders to bring violations to an end, or to seek any other form of remedy or redress. The Proposal instead focuses on the imposition of financial penalties and market sanctions where an AI system is non-compliant. For example, Member States are empowered to implement rules on sanctions, and market surveillance authorities can require that a non-compliant AI system be withdrawn or recalled from the market.[76] Yet, to reach the Proposal's goal of robustly protecting fundamental rights, individuals who encounter AI systems in the EU must be able to rely on a clear system of remedies protecting them from harms and fundamental rights interference caused by public or private adoption of AI systems. As outlined in section 4.1.1, an independent body must then have competence to assess whether any fundamental rights interferences are necessary and proportionate in a democratic society.

Individual rights of redress would address the current gap in enforcement which is especially blatant in cases where individuals cannot reasonably opt out of certain outcomes produced by AI systems (as recognised by the Proposal). Recital 38 calls for 'effective redress' in relation to procedural fundamental rights violations (and draws attention to right to an effective remedy), but classifying something as high-risk is itself *not* a remedy.

Although the risk classification process for AI systems under Article 7(2)(h)(I) includes considering whether processes for remedies are necessary for a particular AI system, there is no guidance on what an effective remedy looks like and no provisions through which individuals can access such a remedy. Additionally, the imposition of penalties for breaches of duties and obligations arising from the Proposal are not individual remedies sufficient for the ongoing robust protection of fundamental rights – rather, they offer only a 'deterrence-based' approach for which there is no guarantee of success, despite clear risks to the fundamental rights of individuals.

It is our understanding that the European Commission will soon publish a draft liability framework for AI systems which could potentially strengthen the procedural rights of individuals who are adversely impacted by AI systems. It is, however, regrettable that this framework was not published simultaneously with the Proposal, since this renders it difficult to holistically assess the protection offered to individuals. We hope that the concerns raised above are either reflected in the revised AI Regulation, or in its accompanying liability rules – or preferably both. The inclusion of individual rights to redress would contribute substantially to an enforcement framework which empowers all relevant stakeholders, and therefore to Legally Trustworthy AI.

b) The Proposal does not provide a complaints mechanism

In addition to a lack of rights which grant individuals standing, the Proposal currently also does not provide the possibility for individuals to file a complaint with the national competent authority – even though the Proposal renders this authority the sole actor who ensures compliance with the Proposal. This absence of a complaints mechanism stands in sharp contrast to the mechanism provided under the GDPR, whereby each supervisory authority has the task to "handle complaints lodged by a data subject, or by a body, organisation or association in accordance with Article 80, and investigate, to the extent appropriate, the subject matter of the

[76] European Commission, "The Proposal," Article 65(5).

complaint and inform the complainant of the progress and the outcome of the investigation within a reasonable period, in particular if further investigation or coordination with another supervisory authority is necessary."[77] Article 77 of the GDPR further provides for a 'right to lodge a complaint with a supervisory authority' for data subjects.

While the Proposal does not preclude national competent authorities from establishing a complaints mechanism on their own initiative, the absence of harmonisation of such initiatives renders it likely that individuals in different EU Member States face different levels of protection. The inclusion of a provision which mandates a complaints mechanism, similar to the GDPR's, would be beneficial to the Proposal's aim of ensuring the protection of fundamental rights. Moreover, it would contribute to the rule of law across the Union. Firstly, it would provide individuals with a clear procedure to follow in case they suspect that an AI system they are subjected to does not meet the requirements of Title III or operates in contravention of Title II of the Proposal. Secondly, it would help national competent authorities to fulfil their tasks, since these complaints can lead to more effective monitoring and evaluation of problematic AI practices.

4.2.4 The Proposal's enforcement mechanism is inadequate

In addition to the absence in the Proposal of the individual affected by AI systems, the Proposal's enforcement architecture copes with several practical problems. For the proposed Regulation to be Legally Trustworthy, it must conform to rule of law standards, and its enforcement must therefore be congruent with the promulgated norms. This may be undermined by the fact that the current enforcement structure is relatively complex and heavily relies on the competencies of national authorities, which may be uneven and under-resourced (a). Furthermore, the role of notified bodies and the scope of the procedural rights conferred on the Proposal's legal subjects could be further clarified (b).

a) The enforcement structure hinges too much on national competencies

The enforcement of the Proposal risks being undermined by the practical aspects of Member State competencies. Article 59 of the Proposal enables Member States to designate or establish national competent authorities "for the purpose of ensuring the application and implementation of this regulation."[78] Unless otherwise provided for by the Member State in question, these authorities will be required to act as *both* 'notifying authority' and 'market surveillance authority' as part of a combined 'national supervisory authority.'[79] Notified bodies are responsible for verifying the conformity of high-risk systems, and market surveillance authorities are responsible for the evaluation of high-risk AI systems "in respect of its compliance with all the requirements and obligations" of the Proposal.[80] If a given system is not compliant, the market surveillance authority shall take "all appropriate provisional measures to prohibit or restrict the AI system's being made available on its national market, to withdraw the product from that market or to recall it."[81] There are three practical concerns about this setup.

---

77 See European Parliament and Council, "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, 4.5.2016", 2016, 1–88 ("GDPR"), Article 57.

78 European Commission, "The Proposal," Article 59(1).

79 European Commission, "The Proposal," Article 59(2).

80 European Commission, "The Proposal," Article 65(2)

81 European Commission, "The Proposal," Article 65(5).

Firstly, this envisaged enforcement architecture remains a work in progress. As it stands, much work is required to determine the contours of the enforcement practices which these bodies must undertake to ensure full compliance with the Proposal. It also raises questions about how these bodies should cooperate with existing entities, from data protection authorities to (notifying) bodies active in the domains and sectors listed in Annex II of the Proposal.

Secondly, there is a risk of uneven implementation across Member States. The Proposal places the burden on national competent authorities to effectively design the proposed regulatory mechanisms at a domestic level. However, the uneven availability of resources in Member States, and different manners in which national authorities could concretise the implementation of the Proposal, risk impeding the development of a common level playing field of fundamental rights protection vis-a-vis AI-generated risks across the EU. This also extends to the development of rules in relation to penalties for non-compliance.[82] It is the role of the European Artificial Intelligence Board[83] to harmonise different approaches to the Proposal's implementation, yet it remains to be seen how this Board will operate in practice.

Finally, the Proposal must acknowledge concerns raised in the context of the GDPR relating to over-reliance for enforcement at the domestic level. Three years after the GDPR's implementation, it has become evident that the protection is significantly weakened due to lack of resources across the EU.[84] It is important not to replicate this mistake with the Proposal. Any enforcement architecture implemented as a result of the Proposal should be developed with careful attention, ensuring that it is (a) appropriately resourced, and (b) that these bodies are staffed with people who have the appropriate skills and knowledge with regard to both technical aspects of AI *and* the social and legal aspects of fundamental rights. Reliance on Member States to complete the work of developing the Proposal's enforcement architecture may be inadequate if not backed up by sufficient financial and human resources, and harmonised implementation guidance.

b) The enforcement powers conferred upon supervisory authorities should be clarified

There are also a number of outstanding issues in relation to specific powers of enforcement conferred upon supervisory authorities.

Firstly, the powers of the supervisory authorities provided for by the Proposal focus on *ex post* controls, yet it remains to be seen on which basis and at which scale these controls will be carried out. The EU database of stand-alone high-risk AI systems, in which those systems need to be publicly registered, could provide a helpful tool to enable supervisory authorities to prioritise their inevitably limited resources. Nevertheless, a risk of under-enforcement remains, particularly given the lack of a complaints mechanism that allows individuals to flag potentially problematic AI practices, as mentioned above.

Secondly, Member States can deviate from the protection afforded by the Proposal if they find there is ground to do so: "*for exceptional reasons, including public security, the protection of life and health of persons, environmental protection and the protection of key industrial and infrastructural assets*."[85] This exception is broad in scope and should be clarified further to

---

82 European Commission, "The Proposal," Article 71(1).

83 This Board, established by Article 56 of the Proposal, shall be chaired by the European Commission and composed of the national supervisory authorities (represented by the head or equivalent high-level official of that authority) and the European Data Protection Supervisor.

84 Estelle Massé, "Two years under the EU GDPR: An implementation progress report - state of play, analysis and recommendations," *Access Now*, May, 2020, https://www.accessnow.org/cms/assets/uploads/2020/05/Two-Years-Under-GDPR.pdf.

85 European Commission, "The Proposal," Article 47.

prevent unjustified infringement of fundamental rights by systems which are otherwise considered high-risk. This is particularly important given that several high-risk AI systems are operated by the authorities of the Member State itself, and granting Member States the power to derogate from the protection offered by the Proposal with regard to their own AI systems risks incentivising abuses of power.

Finally, due attention should be given to the procedural rights conferred on AI providers. In the course of their activities, supervisory authorities have wide access to data and documentation relating to high-risk AI systems. While this is needed to ensure that the requirements can be checked and enforced, the Proposal's procedural limits to these powers are currently relatively scarce. In providing the right of the notified body to have full access to training, certification and test data and to request access to source codes, the Proposal generates a tension between (a) the need to regulate the activities of organisations responsible for the development of high-risk systems, and (b) the need to protect the intellectual property of said organisations, in line with the freedom to conduct business and the right to protection of intellectual property, which are both protected by the EU Charter.[86] It should be ensured that the know-how of businesses is adequately protected, with sufficient confidentiality requirements, and any access requests should be targeted and proportional to the specific task at hand.

The rule of law, as the second pillar of Legally Trustworthy AI, requires that the legal framework for AI is consistent, coherent, and that it is effectively and legitimately enforced. The previous sections argued that the Proposal falls short in this regard, as it relies too much on the instrument of conformity assessments; does not pay sufficient attention to coherence with fundamental rights law and other EU legal instruments; does not grant individuals any procedural rights in its enforcement mechanism; and still requires guidance as regards the practicalities of the Proposal's enforcement. The next sections address the extent to which the Proposal's to safeguard the third pillar of Legal Trustworthiness: democracy.

4.3 The Proposal fails to ensure meaningful transparency, accountability and rights of public participation (democracy)

As the final pillar of Legal Trustworthiness, democracy requires that members of communities are entitled to participate actively in determining the norms and standards that apply to the life of that community. This pertains particularly to activities which have a direct impact on the rights, interests and legitimate expectations of its members, and the distribution of benefits, burdens and opportunities associated with new technological capacities. The inevitability of normative trade-offs in the design and deployment of many existing and anticipated AI applications makes meaningful participation rights especially important. Put simply: democratic participation in the formation of the legal framework around AI contributes to its trustworthiness. Closely related to this idea of Legal Trustworthiness through democracy is the right for individuals to be informed about decisions that affect them in a substantive manner, and of the way in which the commonly agreed upon standards are accounted for. Indeed, meaningful transparency and accountability are essential requirements of a well-functioning democratic society.

We argue that the current Proposal does not do enough to acknowledge the importance of democracy for Trustworthy AI, nor does it provide the institutional framework to robustly incorporate the value of democracy in AI governance. Public consultation and participation rights are entirely absent from the Regulation (4.3.1). Moreover, as highlighted earlier, the Proposal does not provide individuals with any substantive rights, and only formulates

[86] See European Union, "EU Charter," Articles 16 and 17.

transparency *obligations*, which currently fail to ensure meaningful accountability (4.3.2). Indeed, the obligations envisaged in the Proposal are primarily operationalised via bureaucratic management techniques. Finally, the Proposal's reliance on standard-setting mechanisms to operationalise the technical requirements meant to protect fundamental rights risks further increasing this democratic deficit (4.3.3).

4.3.1 The Proposal does not provide consultation and participation rights

The opportunity for meaningful public deliberation – on the basis of which legislation can be enacted – is especially important for politically controversial matters such as the AI prohibitions or a 'high-risk list.' Yet the Proposal's lack of consultation and participation rights seems to fall short of this requirement. First, it is questionable whether the Proposal's content reflects the concerns that citizens have about prohibited and high-risk AI systems (a). Second, the Proposal does not foresee any participation and consultation rights for stakeholders in the context of future revisions (b).

a) The scope of the public consultation prior to the Proposal's drafting would have benefitted from more targeted questions regarding prohibited and high-risk applications

From the outset, the European Commission aimed at ensuring that the voices of citizens and relevant stakeholders could be heard in order to give shape to the proposed Regulation. Thus, a public consultation was organised on the *draft Ethics Guidelines* of the HLEG (published in December 2018), which served as an inspiration for the Commission's further work on AI, as well as on its subsequently refined Assessment List in the form of a broad piloting phase (until December 2019). Furthermore, public consultations were also launched after the Commission's publication of the *White Paper on Artificial Intelligence* (in February 2020), and on the *Inception Impact Assessment* of the proposed AI Regulation (until September 2020). Moreover, we warmly welcome the opportunity provided by the European Commission to provide our feedback on the published Proposal for an AI Regulation in the context of the current public consultation, running until August 2021.

While these efforts of the Commission should be commended, the scope of the last consultations on the *White Paper* and *Inception Impact Assessment* would have benefitted from more targeted questions regarding the stance on practices that should be prohibited and practices that should be considered as high-risk AI systems. The structure of the current Proposal accords great importance to an AI system's status as either 'no-risk,' 'prohibited,' or 'high-risk.' It is unclear whether the current lists under Title II and Title III duly reflects popular concerns about AI systems – particularly in light of the impact they can have on fundamental rights. We doubt whether the lack of emotion recognition systems and biometric categorisation systems under the list of high-risk AI systems (or even prohibited practices) truly echoes the opinion that European citizens have towards such practices. In any case, it is not clear that democratic deliberation is at the foundation of these distinctions.

b) Insufficient opportunities for consultation and participation enshrined in the Proposal itself

The risk that popular opinion on the harms associated with AI is not reflected in the content of the Proposal is exacerbated by the fact that the Proposal itself does not provide consultation rights for the future revision of its content – despite the European Commission's competence to update the Proposal's various Annexes. The provision of such rights is especially important given that the many current and anticipated AI applications, particularly those which entail the collection and analysis of biometric data, claim to offer myriad 'benefits' without any robust

evidence to justify their highly intrusive interference with the fundamental rights of individuals, in real time and at scale.

While provisions are made for the Commission to consult with 'experts' if it wants to expand the list of high-risk systems, or the requisite technical documentation, this is no substitute for public participation by lay members of the community. The current Proposal fails to recognise that one of the legal system's most important roles is to provide an institutional framework through which a community can discuss and determine its own collective values, and how it decides on inescapable conflicts between individual and collective values, and between collective values inter se. Consider in this regard the insistence of the HLEG, in its Policy and Investment Recommendations, that:

> Questions about the kinds of risks deemed unacceptable must be deliberated and decided upon by the community at large through open, transparent and accountable deliberation, taking into account the EU's legal framework and the obligations under the Charter of Fundamental Rights.[51]

Accordingly, when the Commission revises the list of high-risk AI systems, it is crucial that participation of the public at large is ensured within this revision exercise.

Additionally, the absence of any form of public consultation rights is problematic in relation to the determination of what constitutes an 'acceptable' residual risk in the context of high-risk AI systems, which is treated in the Proposal as a matter which developers and deployers are free to determine themselves. As noted in section 4.1.5.a, any 'residual risk' is currently dealt with by requiring that provider communicates these risks to the user, and that the user complies with the stated instructions. Given that the 'user' is the organisation deploying the AI system, this gives so little protection to individuals and communities that it cannot be properly described as inviting democratic deliberation or as respectful of fundamental rights. More generally, as stressed in section 4.1.5.a, the protection of individuals and the general public against the threats posed by high-risk AI systems offered in the current Proposal relies heavily on the effectiveness and legitimacy of the AI provider's compliance with the proposed mandatory requirements, and falls far short of living up to the aspiration of respecting fundamental rights and safety, which requires, at minimum, open public discussion and consultation in relation to what constitutes an 'acceptable residual risk.'

4.3.2 The Proposal lacks meaningful substantive rights for individuals

The fact that the Proposal fails to provide robust frameworks for public consultation and participation might be a symptom of the wider problem that the Proposal seems to have forgotten about the persons subjected to AI systems, or simply put, 'ordinary people.' The absence of rights for 'ordinary people' is not only potentially detrimental to the rule of law, as argued in section 4.2.3, it also directly contributes to the Proposal's democratic deficit.

The procedural individual right to redress and the individual right to a complaints mechanism argued for in section 4.2.3 (in the context of the Proposal's enforcement) would not only strengthen the Proposal's enforcement mechanism, they would also contribute to the Proposal's democratic legitimacy. This is because granting 'ordinary people' a more substantive role in the enforcement of the Proposal might make it more likely that the concerns of those people reach the authorities and are dealt with accordingly. Yet, the Proposal would further benefit from adding substantive rights for individuals. These substantive rights would make the 'ordinary person' more of a central figure in the proposed Regulation, and would provide very specific grounds for them to exercise the procedural rights argued for in section 4.2.3. Two substantive rights which should be granted to individuals are the right not to be subjected to AI systems that disproportionally affect their fundamental rights (in particular, prohibited AI

practices and high-risk AI systems that do not conform to the requirements set out in this Proposal) (a), and information rights which allow individuals to effectively exercise their existing rights (b).

a) The Proposal does not provide any substantive rights for individuals

The lack of substantive individual rights in the Proposal reduces individuals to entirely passive entities, unacknowledged and unaddressed in the regulatory framework. The Proposal's silence on individuals is especially striking considering that one of the primary reasons why AI is being regulated at all is to protect those very individuals from the risks generated by AI systems.

The absence of procedural rights of redress and a right to file a complaint with a national supervisory authority was already set out in section 4.2.3. In addition, the Proposal does provide individuals with any new substantive rights either. It only consists of obligations imposed on AI providers - and to a lesser extent, on users and other actors. Indeed, as remarked earlier, the Proposal offers no equivalent to the 'data subject rights' which can be found under the GDPR. Consider Chapter III of the GDPR, which offers citizens whose personal data has been processed a set of rights which they can exercise against data controllers. These include a right to be informed about personal data collected, a right of access to information regarding why and how their data is being processed, a right to 'rectification' of e.g., incomplete or incorrect data, a right to be 'forgotten,' and rights relating to automated decision-making, including profiling. In addition, data controllers must facilitate the exercise of these rights, without undue delay, unless they can demonstrate that they cannot identify the data subject in question.

No similar rights are offered in the Proposal, wherein responsibility is left mainly to providers and users to monitor the design, implementation and use of AI systems. One can wonder why the Commission did not see the need for any substantive rights in the context of AI, and whether it considers data subject rights as sufficient to tackle relevant dangers posed by AI to fundamental rights more generally. If the Proposal assumes that data protection law provides sufficient individual rights to protect against the risks of AI systems to individuals, it is not clear why, and the equivalence between AI systems and general 'data processing' must be justified.

Concretely, at the very least, the Proposal should grant individuals a right not to be subjected to prohibited AI practices as listed in Title II, and a right not to be subjected to high-risk AI systems that do not meet the conformity requirements of Title III. This would transform the obligations of AI providers and users from merely obligations towards the regulators of the market to obligations towards specific individuals. Making the individual a central figure in the Regulation would reintroduce the 'ordinary person' both as an explicit beneficiary of the Regulation, and as an empowered legal actor in the regulatory framework around AI.

b) The Proposal does not provide meaningful information rights for individuals

For a democracy to flourish, the public must be adequately informed to be able to participate in the political life of their community and to plan their own individual lives. In this context, it is essential that individuals are provided with sufficient information about technological developments which directly or indirectly affect their health, safety, and fundamental rights. Additionally, democracy requires that the public can challenge those in power to keep a check on their actions, which can also only be done if sufficient information is available to the public. Both these democratic desiderata require a governance framework which affords an

appropriate degree of *transparency* which is itself a prerequisite for *meaningful democratic accountability*.

The Proposal addresses the issue of transparency in three ways, which are yet cumulatively insufficient. Firstly, particular AI systems (like emotion recognition systems) are subject to transparency obligations specified in Title IV. These obligations amount to informing the individuals subjected to such systems that they are, in fact, being subjected to them. As mentioned in section 4.1.4.c, merely telling someone they are being subjected to a system does not amount to protecting that person against the adverse effects of that system and could even cause chilling effects. Secondly, under Article 13 of the Proposal, AI providers have certain information obligations towards AI users as regards the AI system's features. As outlined in section 4.1.5.e, these transparency obligations are hence owed to the user of the AI system, not the individuals exposed to them. Thirdly, under Title VII of the Proposal, a publicly accessible EU database is established for stand-alone high-risk AI systems, yet the information that needs to be provided in that database is fairly limited, and does not provide individuals with sufficient information to potentially question and contest the AI system's impact. In sum, while the Proposal does explicitly address the need for transparency, it does not guarantee that the general public receive sufficient information to understand the risks which they are being subjected to. Moreover, these transparency obligations are not grounded in a framework which gives individuals clear pathways for contesting the existence or operation of certain AI systems and thereby using the obtained information in a way which contributes to fundamental rights protection. Transparency is best used as a method of exposing AI systems to public scrutiny and fundamental rights analysis, as a crucial element of a wider system of rights protection, but it does not equate to protection in and of itself.

Transparency with a view of protecting both democracy and fundamental rights should (1) primarily focus upon the information needed to expose potential risks to, and violations of, fundamental rights; and (2) should thus be tied to clear *actionable* means by which to remedy such occurrences. This means that transparency rights and obligations need to be tied directly to the needs of affected individuals to understand precisely:

(a) How they are affected by the AI system in question, including how the AI system generates and arrives at outputs from a given set of inputs which then directly affect them and in which ways;

(b) How any effects incurred may interact or interfere with their fundamental rights (including the provision of a reasoned explanation for any relevant adverse decisions);

(c) How they may take action and obtain remedies where they are concerned that their fundamental rights have been unduly or disproportionately affected – or where AI providers and users may have otherwise failed to fulfil legal obligations.

The Proposal currently does not deliver these elements, and should therefore be strengthened in the following manner:

First, in line with the recommendation in section 4.1.5, the information obligations currently imposed on AI providers in Title III should be extended to individuals subjected to the AI system, rather than merely targeting its commercial users. This can also be reflected in a substantive right for affected individuals to request such information from the provider or user of the AI system.

Second, the transparency obligations of Title IV imposed on certain AI systems in light of the risk of manipulation they pose, should be complemented with a substantive *right* to be informed about the use of those systems. This should be the case for all three of the AI systems

covered by Title IV, and is particularly important for BCS and ERS, considering the intrusive and risky nature of these systems. The fact that these systems are currently only covered by transparency obligations in Title IV is emblematic of the Proposal's poor recourse to 'transparency obligations' to protect fundamental rights. Merely mandating AI providers to inform people that they are being subjected to intrusive technologies which are unjustified by scientific evidence, and potentially discriminatory, does not address the chilling effects of these technologies, but rather enhances them. Therefore, in addition to including a *right* to be informed about the use of these systems rather than only relying on an information obligation in Title IV, in section 4.1.4 we also argued that BCS and ERS should be subject to the requirements for high-risk AI systems of Title III and, in some instances, fall under the prohibited AI practices of Title II.

Third, the information that should be included in the EU database of stand-alone high-risk AI systems (as per Annex VIII) should be extended. It offers, for instance, little comfort to an individual who suffered unjust interferences from the use of an AI system to detect and prevent 'irregular immigration' (identified in Annex III as a high-risk system), if they can see that this system has been included on the database without having access to information that helps them to understand and interrogate its legality. Such information could include, for instance, the characteristics, capabilities and limitations of performance of the high-risk AI system, including not only its intended purpose, but also its level of accuracy, known or foreseeable circumstances related to the use of the high-risk AI system which may lead to risks to the health and safety or fundamental rights, the underlying assumptions on which the system is operating, the system's performance as regards the persons or groups of persons on which the system is intended to be used and, when appropriate, relevant information about the training, validation and testing data sets used, taking into account the intended purpose of the AI system.[87]

Fourth, to complement the right to a complaints mechanism argued for in section 4.2.3, individuals should be able use the information they receive to challenge and contest the problematic use of AI systems. Decisions resulting from these systems can implicate a range of fundamental rights (in the above example, for instance, the right to liberty and security, the right to conduct business, non-discrimination - and where transparency is lacking - the right to an effective remedy).

In sum, transparency rights and obligations must be embedded in a larger framework of rights to achieve what they claim to achieve. Merely informing an individual that a risky AI system has been used offers little protection if that individual does not have rights to essential information about that AI system, and does not have a procedure through which to use that information to contest the system before an independent authority. If the Proposal can be strengthened to ensure that each of the elements set out above are provided for, this would demonstrate a much stronger commitment to empowering individuals in exercising and safeguarding their fundamental rights, and ensuring accountability for any adverse impacts caused by the use of AI systems. Not only would this afford a more central role for 'ordinary people' in the regulatory framework and therefore contribute to the third pillar of Legal Trustworthiness, it would also strengthen the protection of fundamental rights and the rule of law.

[87] Inspiration can *inter alia* be found when considering the transparency obligations under Article 13 of the Proposal, which are currently solely directed to the commercial user of the AI system.

4.3.3 The Proposal suffers from a democratic deficit in standard-setting and conformity assessment

The final aspect of the proposed Regulation which presents a potential threat to the third pillar of Legal Trustworthiness is the process by which the implementing standards and conformity assessments are shaped.

As noted above, the regulatory framework laid down by the Proposal embraces the EU's 'New Approach' for goods. This means that the general requirements for high-risk AI systems are laid down in the Proposal (the so-called 'essential requirements'), while the detailed technical requirements will be set out primarily in European standards developed through European standardisation. Even though the detailed technical standards have already been ascribed a large role in Title III's Chapter 5, they are largely still lacking today. Their development will be crucial for the effective implementation and enforcement of the proposed Regulation.

However, as already noted by Veale and Zuiderveen Borgesius, while standardisation bodies such as CEN/CENELEC can be empowered by the Commission to develop the standards, this approach is not without controversy.[88] Standards creation in principle constitutes an open process, yet consumer representatives and civil society organisations are typically underrepresented therein due to a lack of resources and domain knowledge. Coupled with the more general absence of public participation mechanisms in the Proposal as outlined above, this approach risks leading to a process which provides insufficient democratic input in a process which significantly affects the practical application of the proposed Regulation. It will hence be crucial that the Commission take measures to ensure that not only private entities, but also organisations which represent public interests are involved in the creation of standards which might be technical, but value-laden and contested nonetheless.

This point can also be made more generally regarding the implementation of the Proposal's conformity assessment mechanism. The conformity assessment of AI systems is performed according to technical rules which are entirely defined by notified bodies – i.e. private bodies which in principle receive fees for their activities. It is hence of utmost importance to ensure that – to the extent possible – national authorities are nevertheless empowered to exercise democratic control on how these entities perform their activities, and on how the Proposal's standards are implemented concretely.

To conclude, the third pillar of Legal Trustworthiness requires that the regulatory framework around AI provides ample opportunities for citizens to influence the priorities of the legislation, its regulatory design, and the details regarding its implementation. This requires that the 'ordinary person' is given a central role in the proposed Regulation. Accordingly, the Proposal could be significantly strengthened if it provided explicit procedures for public participation and consultation; meaningful substantive rights for individuals; meaningful transparency obligations embedded in a larger framework of fundamental rights protection; and procedures which encourage and enable democratic input for the setting of technical standards.

5. OUR KEY RECOMMENDATIONS

In this document, we contextualised the proposed AI Regulation against the background of previous efforts by the EU to develop a framework for Trustworthy AI. We identified the attainment of 'Legally Trustworthy AI' as the main goal of the Proposal, in order to fill the gap of the High-Level Expert Group's Ethics Guidelines which explicitly only dealt with 'Ethical AI' and socio-technically 'Robust AI.' We argued that the concept of 'Legal Trustworthiness' consists of three pillars – fundamental rights, rule of law, and democracy – and showed how

[88] Veale & Zuiderveen Borgesius, "Demystifying the Draft."

these can be attained (i.e., through appropriate allocation of responsibility, adequate enforcement mechanisms and coherence, and processes which encourage public participation). In the sections detailing how the Proposal falls short of these three pillars, we made numerous recommendations for the Commission to consider. For the sake of convenience, our explicit recommendations are listed below, organised by the Title and Article they pertain to.

5.1 Recommendations for Title I of the Proposal

Article 2 and Article 3 (scope and definitions) – as discussed in section 4.1.2:

- Consider **broadening the scope** of the Proposal to explicitly include all computational systems used in the identified high-risk domains, regardless of whether they are considered to be 'AI.' This might require some revision of the logging requirements to make them more inclusive of systems which do not necessarily rely heavily on data. Such broadening would make the application of the Proposal more dependent on the *domain* in which the technology is used (as specified in Annex III) and the fundamental rights-related risks related thereto, rather than on the specific computational *technique* used to engender the risks.
- Clarify the position of university **researchers undertaking academic research** under the Proposal. Ensure equal levels of fundamental rights protection for research done by university researchers in academic institutions and in corporate R&D departments. Also explicitly include in-the-wild trials or experiments under the definition of "putting into service" of Article 3(11).
- Remove the condition "where that use falls under the exclusive remit of the Common Foreign and Security Policy regulated under Title V of the Treaty on the European Union (TEU)" from recital 12, in order to prevent the establishment of different legal regimes for **military AI** systems depending on the context in which they were developed – OR repeat that condition in article 2(3), meaning that all military AI systems developed outside of the CFSP would fall under the scope of the Proposal, and include military AI as a high-risk category. If all military AI is to be excluded from the scope of the Proposal, this leaves a gap in legal protection against the large risks associated with military AI. Ensure that military AI is regulated to the extent that the Union has legal competence (at least the use of military AI in the context of the CFSP), if necessary, under a different legal basis than the current Proposal.
- Include the use of AI systems for **national security** and intelligence agencies as a high-risk domain and expand Article 3(40) to include these agencies in order to prevent a loophole undermining the prohibition on real-time remote biometric identification.

5.2 Recommendations for Title II of the Proposal

Article 5 (prohibited practices) – as discussed in section 4.1.3

- Add a procedure which enables the Commission to **add prohibited practices** to Article 5 of the Proposal after review and consultation. Include a set of criteria which the Commission should use to determine whether a particular AI practice should be prohibited, similar to the list in Article 7(2) for high-risk systems, to provide legal certainty. Ensure that the list of high-risk systems does not include AI practices which are incompatible with fundamental rights and whose use cannot be reasonably justified, such as technologies which are manifestly discriminatory and/or lacking in any clearly established scientific basis. Consider moving such AI practices to the list of prohibited practices.

- Expand the scope of Article 5(1)(a) and (b) to **include harms other than physical and psychological** harm. Remove references to physical and psychological harm and replace them with 'harm' and 'fundamental rights interference,' and consider expanding the scope to also include harms to groups and to EU values as listed under Article 2 TEU.
- Expand the scope of Article 5(1)(a) to include **manipulative AI practices** which do not rely on subliminal cues, but which nevertheless cause harm or interfere with fundamental rights.
- Expand the references to **vulnerable groups** in Article 5(1)(a), (b) and (c) to include all protected characteristics listed in Article 21 of the Charter of Fundamental Rights of the European Union.
- Consider expanding the scope of the prohibition on **social scoring** in Article 5(1)(c) to private actors which operate in social domains with significant effects on individuals' lives.
- Add a reference in Article 5(1)(c) to **proxies** for personal characteristics used for social scoring practices.
- Expand the prohibition on remote biometric identification systems in public spaces of Article 5(1)(d) to **non-law enforcement public actors**, and at the very least to actors with coercive powers. Consider extending this prohibition to private organisations, or *at least*, subject these organisations to obligations that are more robust and significant than the current high-risk requirements.
- Consider the effects of the exceptions to the prohibition on law enforcement use of remote biometric identification systems for the construction of technological **infrastructures** which in principle allow for such practices. These infrastructures could cause chilling effects on fundamental rights without even being used, and could be subject to function creep.
- Prohibit both the use of remote live **biometric categorisation** systems in public places and the use of **emotion recognition** systems, by law enforcement and other public actors with coercive power.
- Add biometric categorisation systems and emotion recognition systems that would not fall under this suggested prohibition to the list of **high-risk systems**. Consider subjecting these systems to independent *ex ante* control, especially when they are used in safety-critical and fundamental rights-critical domains, on vulnerable individuals and groups, and in situations with power asymmetries.

5.3 Recommendations for Title III of the Proposal

Article 6 (classification of high-risk systems) – as discussed in section 4.1.5

- Include **law enforcement** AI systems which do not explicitly use personal data (e.g. crime hotspot analysis based on geospatial data) in the list of high-risk systems in Annex III(6).
- Consider extending the 'list-based' approach with an approach based on **broader risk criteria**, as can also be found under the GDPR.
- Consider ***ex ante*** **verification by an independent body** for a broader set of high-risk AI systems, beyond just biometric systems – especially when used in the context of

vulnerable groups or when based on unscientific approaches such as in the case of emotion recognition or polygraphs.

Article 7 (amendments to Annex III) – as discussed in section 4.1.5

- Empower the Commission to add new AI points (in addition to points 1-8) to the list of high-risk AI systems. Ensure that the **adding of high-risk categories** is done through robust consultation with stakeholders, including citizens and public interest representatives.

Article 10 (data and data governance) – as discussed in section 4.1.5.e

- Add to Article 10(3) a requirement concerning **data provenance and data integrity**, to enable checks on the legitimacy of the origins of the data.

Article 13 (transparency and provision of information) – as discussed in section 4.1.5.e and section 4.3.2

- Add to Article 13(3) the requirement to provide information about the ways in which those subjected to the system may be **adversely impacted** by the system. Include a statement about the **conditions 'in the field'** under which the AI system is intended to be used, as well as the parameters of performance testing.
- Include individuals subjected to AI systems as beneficiaries of the information obligations imposed on AI providers in Title III. Reflect this inclusion in a substantive **information right** for affected individuals to request the specified information from either the provider or user of the AI system.

Article 14 (human oversight) – as discussed in section 4.1.5.e

- Add to Article 14(3) an organisational, **non-technical category of human oversight** measure which consists of at least: training for decision-makers, logging requirements, and clear processes for *ex post* review and redress. Ensure that human oversight does not legitimise rights-violating uses of AI.

Article 16 (obligations for providers of high-risk systems) and Article 29 (obligations for users of high-risk systems) – as discussed in section 4.2.2.c

- Clarify the relationship between data protection impact assessments under the **Law Enforcement Directive** and impact and conformity assessments under the Proposal. Additionally, clarify the duties of private AI providers in the context of law enforcement activities in light of Article 3(7)(b) of the LED.

Article 43 (conformity assessment for high-risk systems) – as discussed in section 4.1.5.a

- Expand the list of high-risk systems which are subject to **prior independent conformity** assessment control. This should particularly be considered for AI systems that are used in contexts of asymmetry of power (such as, for instance, migration management and law enforcement), systems used for the biometric categorisation of individuals (which are currently not listed in Annex III), and systems relying on unscientific methods (such as polygraphs and emotion recognition systems, regardless of their deployment by a private or public actor).
- Ensure that conformity assessment does not amount to mere 'tick-box' exercises. Incorporate a requirement for the AI provider to engage in a **discourse of justification** regarding potential fundamental rights interferences, requiring that providers engage

robustly with proportionality and necessity assessment, in plain language and available for independent review.

Article 47 (derogation from conformity assessment procedure) – as discussed in section 4.2.4.b

- Clarify the circumstances under which Member States may **derogate** from the conformity assessment procedure. In doing so, ensure that this provision prevents abuses of power in cases where Member States themselves wish to deploy high-risk systems.

5.4 Recommendations for Title IV of the Proposal

Article 52 (transparency obligations) – as discussed in section 4.1.3.b

- Subject AI systems which rely on **subliminal cues** without causing any harm or fundamental rights interferences to the transparency requirements of Title IV.
- Add an explicit **right for individuals** who are subjected to the AI systems falling under transparency obligations of Title IV **to be informed** about the use of such systems.

5.5 Recommendations for Titles VII and VIII of the Proposal

Article 60 (EU database for stand-alone high-risk AI systems) – as discussed in section 4.1.3.b

- Consider **expanding the amount of information** required to be provided by the AI provider in the context of the EU database (as per Annex VIII), to grant individuals access to information which helps them interrogate the legality of the high-risk systems they are subjected to. Such information could include the characteristics, capabilities and limitations of performance of the high-risk AI system, known or foreseeable circumstances related to the use of the high-risk AI system which may lead to risks to the health and safety or fundamental rights, the underlying assumptions based on which the system is operating, the system's performance as regards the persons or groups of persons on which the system is intended to be used and, when appropriate, relevant information about training, validation and testing data sets used.

Title VIII (post-market monitoring) – as discussed in sections 4.2.3.b, 4.2.4.a and 4.2.4.b

- Add a provision which mandates a **complaints mechanism** before the national supervisory authority for individuals who suspect that an AI system they are subjected to does not meet the requirements of Title III or operates in contravention of Title II of the Regulation.
- Define the **procedural limits** of the supervisory authorities in a manner which upholds intellectual property rights and protects AI providers from undue interference by national authorities.
- Ensure that Member States authorities have access to sufficient **financial and human resources** to effectively implement and enforce the proposed Regulation and provide harmonised implementation guidance to Members States, to prevent uneven or unreliable enforcement at the national level.

5.6 Other Fundamental Rights Recommendations, including redress and participation

Justificatory discourse for fundamental rights interference – as discussed in section 4.1.1.a

- Align the binding content of the Proposal with existing fundamental rights legislation and practice which establishes substantive and procedural requirements for potential

interferences with fundamental rights, including the **principles of proportionality, necessity, and independent oversight**.

- Explicitly impose an obligation on **private sector actors** to respect fundamental rights when AI systems are put into service in a direct and unmediated way, and not merely as legal obligations that imposed upon states.

Rights for individuals – as discussed in section 4.2.3 and 4.3.1

- Clarify the relationship between the **rights afforded to individuals under the GDPR** and the Proposal. If the Proposal assumes that data protection law provides sufficient individual rights to protect against the risks of AI systems to individuals, the equivalence between AI systems and general 'data processing' must be justified.
- Add an explicit **right for individuals not to be subjected** to the prohibited AI practices listed in Title II, and the right not to be subjected to high-risk AI systems which do not meet the conformity requirements of Title III.
- Add an explicit **right of redress** for individuals who are subjected to non-compliant AI systems, similar to the rights of data subjects under data protection law. Ensure that the Regulation, combined with the accompanying liability rules, provides adequate remedies for individuals where the exercise of fundamental rights is implicated by the development or use of AI systems.
- Add an explicit **information right** for individuals who are subjected to high-risk AI systems to be granted the information provided to the users of the AI systems under Title III.
- Add public **participation rights** for EU citizens regarding the decision to amend the list of high-risk systems in Annex III.
- Add public participation rights for those subjected to high-risk AI systems to be involved in the determination of the acceptability of the '**residual risk**.'
- Ensure that not only corporate and 'expert' groups are involved in the setting of technical **standards** for conformity assessments of high-risk AI systems, by actively involving organisations which represent public interests.