Policy briefing

31 March 2022

People, risk and the unique requirements of AI

18 recommendations to strengthen the EU AI Act

The primary purpose of this briefing is to provide specific recommendations for EU policymakers for changes to be implemented into the final version of the AI Act. These proposals are intended to strengthen the legislation and better reflect the objectives of the EU as described in the proposal: regulating AI technologies in ways that align with EU values and ensuring a fertile environment for innovation – while also supporting solutions to some of the most difficult questions about AI regulation. The briefing will also be of interest to global policymakers with an interest in emerging AI regulation.

The Ada Lovelace Institute is a UK and Brussels-based research institute with a mission to make data and AI work for people and society. We do this by building evidence, drawing from and contributing to informed work in the ecosystem, and fostering rigorous debate on how data and AI affect people and society. One of our priorities is to ensure legislation of AI and data-driven technologies reflects the needs of those affected, while providing a strong platform on which other regulatory solutions may be constructed in the future.

Ada Lovelace Institute

For more information about the Ada Lovelace Institute and our work on the EU AI Act, contact Alex Circiumaru: acirciumaru@adalovelaceinstitute.org.

Introduction

In April 2021, the European Commission published its Proposal for the Artificial Intelligence Act ('EU AI Act' or 'the proposed Act'), marking an important moment in the global pursuit to regulate these technologies.

While the European Union ('EU') is not the only global actor seeking answers for the tough questions inherent to the regulation of AI, the progress it has made over the past few years is significant. The AI Act, once adopted, will be the first comprehensive AI regulatory framework in the world.

This 'first-mover' advantage builds on the General Data Protection Regulation (GDPR)'s setting of a global standard for the protection of personal data and means this EU approach to the regulation of AI is likely to have significant global implications.

As the first comprehensive attempt at an AI regulatory framework in the world, the proposed EU AI Act is a significant landmark in the development of global regulation. We recognise the European Commission's leadership in seeking to ensure the development and deployment of 'trustworthy AI' systems.

> The Act aims to create the right regulatory structure to enable AI technology to fulfil its significant potential to improve individual human lives and society, while ensuring the protection of fundamental rights and European values.

Based in the UK and Brussels, the Ada Lovelace Institute maintains an international outlook that recognises the importance of the EU's developing regulatory proposals, both within the EU and globally.

Our work has brought together evidence-based research with expert convenings to influence policy and practice. Particularly relevant to the EU AI Act is our research on public attitudes and legal review of biometrics,[1] technical methods for regulatory inspection of algorithmic systems[2] and algorithmic accountability in the public sector.[3]

---

1 Ada Lovelace Institute. (2021). *The Citizens' Biometrics Council*. Available at: https://www.adalovelaceinstitute.org/report/citizens-biometrics-council/

2 Ada Lovelace Institute. (2021). *Technical methods for regulatory inspection of algorithmic systems.* Available at: https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection/

3 Ada Lovelace Institute, AI Now, Open Government Partnership. (2021). *Algorithmic Accountability for the Public Sector*. Available at: https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/

This policy briefing builds on the expert opinion paper commissioned from Professor Lilian Edwards, a leading academic in the field of internet law, which addresses substantial questions about AI regulation in Europe, looking towards a global standard.[4] This work has been informed by the thinking and contributions of other civil society organisations, academics and specialists.[5]

The 18 recommendations have been refined through research and convening by the Ada Lovelace Institute and revolve around three areas:

1. Ensuring that those who ultimately use or are affected by AI ('affected persons') are empowered to participate in its regulation, from the very first stages – such as standard-setting – through to enforcement.

2. Reshaping the meaning of 'risk', and extending it beyond individual fundamental rights, health and safety, to include systemic and environmental risks.

3. Clarifying and strengthening the governance framework to accurately reflect how AI systems are developed and adapted between different actors.

> We view the recognition of 'affected persons' as part of the framework (recommendation 1) and the creation of a comprehensive remedies framework (recommendation 16) to be the priority for European legislators.

The structure of this briefing follows that of the proposed Act, and recommendations are summarised against the relevant headings. We will continue working on the issues presented here through research and convening, and plan to build on this work with tailored recommendations following consultation with policymakers and other relevant stakeholders in the legislative process.

4 Beruzzi, L. and Noyan, O. (2021). 'Commission yearns for setting the global standard on artificial intelligence.' *Euroactiv*. Available at: https://www.euractiv.com/section/digital/news/commission-yearns-for-setting-the-global-standard-on-artificial-intelligence/. See also: European Commission. 'A European Approach to Artificial Intelligence'. Available at: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

5 See, in particular, this joint statement from more than 100 organisations: https://edri.org/our-work/civil-society-calls-on-the-eu-to-put-fundamental-rights-first-in-the-ai-act/

Summary of recommendations

Title I: Scope and definitions

This section of the Act sets out the envisaged regulatory framework, outlining the recognised actors and relations between them.

The recommendations are designed to accurately reflect how AI systems are developed and adapted, including explicitly building those affected by AI into the framework; to clarify and strengthen the responsibilities of different actors; and to develop the approach to risk categorisation.

1. Add 'affected persons' to the envisaged framework, defined as the natural or legal persons who are ultimately affected by the deployment of an AI system.

2. Rename 'users' to 'deployers', to bring a clearer distinction between those who deploy systems created by providers and those who use or are ultimately affected by the use of AI systems.

3. Determine the categories of risk based on the 'reasonably foreseeable purpose of an AI system' rather than on its 'intended purpose' as defined by the provider.

4. Change the 'substantial modification' test to clarify the dynamic between providers and users, and to ensure legal certainty.

5. Establish clear, judicially reviewable criteria for placing AI systems into categories of risk, building on those already provided in Article 7 of the proposed Act, and including systemic and environmental risks.

Title II: Unacceptable risks and prohibited AI practices

This section of the Act recognises that the use of certain AI systems would give rise to unacceptable risks.

The recommendations are designed to meet the need to provide clear, reasoned categories of risk for different systems, including instances where the use of these systems could be justified despite the risks posed.

6. Considering systemic risks as well as risks to fundamental rights, expand the scope of Article 5(1)(d) to cover retrospective identification, private and online spaces and private actors.

7. Add biometric categorisation and emotion recognition to the scope of Article 5.

8. Make the deployment of any AI system that poses an 'unacceptable risk' possible only in exceptional circumstances where it passes a 'reinforced proportionality test'.

9. Include an obligation to publish all decisions that approve the placing on the market or deployment of unacceptable risk systems in exceptional circumstances.

10. Ensure Articles 5(1)(a) and 5(1)(b) offer meaningful and substantial protection, by removing the reference to 'subliminal', broadening the scope of 'vulnerability' beyond age and disability and adding 'economic damage'.

Title III: High-risk AI systems

This section of the Act considers the definition and potential impacts of high-risk AI systems.

The recommendations are designed to ensure the proposed framework is future-proof and calibrated to take into account all relevant risks, and imposes appropriate, effective, requirements on high-risk systems.

11. Create a mechanism to allow the Commission to add new categories of high-risk AI systems to Annex III.

12. Add systemic and environmental risk to the list of criteria in Article 7.

13. Add a requirement for all high-risk AI systems to be subjected to regular 'impact evaluations'.

14. Add a requirement to ensure that high-risk AI systems are transparent for affected persons, in addition to the provision for an EU database for stand-alone high-risk AI systems in Article 60.

Title IV: Transparency obligations for specific AI systems

This section of the Act is designed to address the potential harms posed by three specific AI systems, but in our analysis does not on deliver the objectives of the proposal or contribute to mitigating the potential risks posed by these three systems.

The recommendation proposes integration into other parts of the Act, to better meet relevant goals and respond to potential risks.

15. Remove Title IV completely, as the systems it addresses fall either in the 'unacceptable' or 'high-risk' category, according to the list of criteria defined in Recommendation 4.

Titles VI, VII, and VIII: Governance and implementation

This section of the Act addresses issues of governance and implementation.

The recommendations support including those who are affected by these technologies in the governance and enforcement of the future Act. This is essential in ensuring the objectives of the Act are met.

16. Create a comprehensive remedies framework for affected persons, including a right for individuals to bring complaints; a right to bring collective action; and a right to information, supplementing what is already provided by the General Data Protection Regulation ('GDPR').

17. Build civil society representation into the mandated standard-setting process and place an obligation on the Commission to review standards before they can be used.

18. Ensure the framework will provide Market Surveillance Authorities with the resources necessary to carry out their role and responsibilities.

Detailed recommendations

Title I: Scope and definitions

1. Add 'affected persons' to the envisaged framework, defined as the natural or legal person who are ultimately affected by the deployment of an AI system

The current framework envisaged by the Act reflects the language of the Commission's product safety approach and recognises only two actors: 'providers' (the entity putting an AI system on the market) and 'users' (those under whose authority an AI system is deployed). This approach fails to account for those who are ultimately affected by the deployment of an AI system, the 'affected persons'.

To accurately reflect the AI lifecycle and include those who are ultimately affected by the deployment of an AI system within the framework laid down by the Act, we propose 'affected persons', which covers those affected even in cases where their personal data is not being used (therefore different to 'data subjects') and they are not actively using the AI system (therefore different to the Act's existing definition of 'users'). For example, students whose final grades are determined by an AI-based system would be affected persons, so would be job applicants whose CVs are processed by an AI-based system. This recommendation will have consequences throughout the Act, and particularly supports recommendation 16, which builds affected persons into governance and enforcement.

Action: The definition of 'affected persons' should be added in Article 3, as a new paragraph, 3(5), and should read: 'the natural or legal person who is ultimately, directly or indirectly, impacted by the deployment of an AI system'.

2. Rename 'users' to 'deployers' to bring a clearer distinction between those who deploy systems created by providers and those who use or are ultimately affected by the use of AI systems

The Commission's current definition of the entity under whose authority an AI system is used as a 'user' rather than a 'deployer' could give rise to confusion and undermine the importance of individuals who are ultimately affected, the end users or 'affected persons'.

The term 'user' should be replaced with 'deployer' to avoid confusion and make the roles of each actor in the AI lifecycle clear.

Action: Article 3(4) should be amended as follows: 'deployer' means any natural or legal person, public authority, agency or other body deploying an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity.

3. Determine the categories of risk based on the 'reasonably foreseeable purpose of an AI system' rather than on its 'intended purpose' as defined by the provider

A feature of the Act is its focus on the 'intended purpose' of an AI system. The intended purpose determines the category of risk and substantially modifying the intended purpose means an identified 'deployer' (formerly 'user', see recommendation 2) becomes a provider.

Article 3(12) defines intended purpose as 'the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation'.

It is therefore for the providers to lay down the intended purpose of a system. But, given the nature and complexity of AI systems, this may not offer adequate clarity about when a deployer has moved beyond intended purpose.

It is essential, particularly for general AI systems, that providers consider the breadth of potential purposes as AI is developed. A clear example of why this is necessary is that a system for 'pattern recognition' could be intended to identify different types of bread and pastries as well as identify cancerous cells.[6]

Changing the language to 'reasonably foreseeable purpose' would require providers to consider more fully the range of potential uses for their technology. It would also encourage greater clarity in setting the limits of the systems providers put on the market as to how far deployers can experiment with an AI system without incurring extra obligations. We are undertaking further work to consider amendments required to ensure the EU AI Act is fit for purpose with regards to general-purpose AI (see ['Regulating AI in Europe: four problems and four solutions[7]] for further discussion).

6 Consumer Technology Association. (2021). 'The AI Pastry Scanner that is now fighting cancer'. Available at: https://www.ces.tech/Articles/2021/May/The-AI-Pastry-Scanner-That-Is-Now-Fighting-Cancer.aspx

7 Edwards, L. (2022). *Regulating AI in Europe: four problems and four solutions*. Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/report/regulating-ai-in-europe/

Action: Change the focus from 'intended purpose' to 'reasonably foreseeable purpose' by amending Article 3(12).

4. **Change the 'substantial modification' test to clarify the dynamic between providers and users and ensure legal certainty**

Changing the focus from intended purpose to reasonably foreseeable purpose would have a positive impact on the substantial modification test, making the meaning of 'substantial' clearer. If a system is used beyond what was reasonably foreseeable that would amount to a substantial modification and to a change in the dynamics between the original provider and the deployer.

This approach will be beneficial for both providers and deployers, who will have a clearer sense of the extent of their responsibilities. Deployers will be given more clarity and therefore confidence to innovate, without however giving them free reign to use a system indefinitely without further checks and balances.

Action: Change Article 3(23) to specify that the use of an AI system beyond the reasonably foreseeable purpose amounts to a substantial modification.

5. **Establish clear, judicially reviewable criteria for placing AI systems into categories of risk, building on those already provided in Article 7 of the proposed Act, and including systemic and environmental risks**
   The proposed Act provides a list of prohibited AI systems (Article 5), one of high-risk systems (composed of Annexes II and III) and one for limited risk systems (Article 52). It is not clear what criteria have been used in placing different AI systems into different categories. Some examples of AI appear to have been miscategorised, for example deep fakes and emotion recognition are placed in the limited risk category, despite the systemic risks they pose – from misinformation to gender hate.

Criteria for risk categorisation would clarify the rationale behind categorisation, and should itself be open to scrutiny and challenge. This would strengthen confidence in categorisation of current AI systems, and further future-proof the Act to enable it to place new uses rationally in appropriate categories.

Action: Amend the proposal to include clear, judicially reviewable criteria for AI systems to be placed in the different categories of risk, developing the foundations laid down in Article 7. The Commission should set out how the criteria have been applied to each individual system. We anticipate criteria would show that AI systems such as emotion recognition or deep fakes should be recategorised.

Title II: Unacceptable risks and prohibited AI practices

6. Considering systemic risks as well as risks to fundamental rights, expand the scope of Article 5(1)(d) to cover retrospective identification, private and online spaces and private actors

Article 5 of the Act is dedicated to prohibited AI practices. It includes real-time, remote biometric identification in public spaces. This narrow list of biometric technologies that are deemed to present unacceptable risks fails to assess adequately the risks use of these technologies poses to both fundamental rights and societal values.

Biometric identification technologies have negative impacts on fundamental rights in circumstances beyond real-time, remote identification.[8] The Act's focus acknowledges the real concerns about the use of covert biometric identification in public spaces by governments and police, which has dominated discourse in recent years. However, the distinction between real-time and 'post' use doesn't protect against concerns about normalisation of surveillance creating chilling effects on freedom of expression and assembly, for example. Indeed, some of the most controversial facial recognition technologies, such as Clearview AI,[9] would qualify as 'post' use.

While there is substantially less research on private uses of biometrics, we anticipate comparable risks to individuals, groups and society, with potentially less justification and access to fewer available remedies, at least from a fundamental rights perspective, than when these technologies are used by public institutions. Leaving private uses out of the scope could lead to perverse results that significantly transform societies and shift the balance of power between people, the

8 See for a more comprehensive discussion: Kind, C. (2021). *Containing the canary in the AI coalmine – the EU's efforts to regulate biometrics*. Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/blog/canary-ai-coalmine-eu-regulate-biometrics/

9 Clearview AI, which describes itself as the world's largest facial recognition network, enables law enforcement to match unknown people to online images scrapped from the internet but has come under criticism for eroding privacy and faced fines from regulators for data gathering without consent. See: Hill, K. (2021). 'The Secretive Company That Might End Privacy as We Know It'. *The New York Times*. Available at: https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

state and private actors, therefore ultimately affecting the rule of law, democratic principles and the values of European societies as outlined by Article 2 of the Treaty on the European Union.[10]

Action: Extend the scope of Article 5(1)(d) to include retrospective biometric identification, carried out in private spaces and by private actors.

7. Add biometric categorisation and emotion recognition to the scope of Article 5

Building on recommendations 5 and 6, we propose that biometric categorisation and emotion recognition pose risks equivalent to real-time, remote identification, to the values of European societies and fundamental rights, and that the scope of Article 5 should be extended to include them.

The use of biometric technologies for classification and categorisation is distinct from identification and presents different concerns, because the underlying categories may be unstable or socially contingent.[11] At their most extreme, these systems use social stereotypes as a basis for drawing correlations, such as examples of systems designed to predict sexuality[12] or criminality[13] from pictures of people's faces. Their use could lead to discrimination on the basis of characteristics that are protected under EU law.

In addition, the link between categories assessed and the collected biometric data points are often not scientifically proven or accurate. Emotion recognition, for example, implies that biometric technologies like facial expression can infer the inner emotional state or internal competencies of a subject. There is no current scientific basis for this technology, and doubts in psychology research as to whether it is even possible to infer emotional states externally.[14] While potential benefits of

10 Article 2 of the Treaty on the European Union reads 'The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail.' Available at: https://eur-lex.europa.eu/eli/treaty/teu_2012/art_2/oj

11 Wendenhorst, C. and Duller, Y. (2021). *Biometric Recognition and Behavioural Detection: Assessing the ethical aspects of biometric recognition and behavioural detection techniques with a focus on their current and future use in public spaces*. Study Requested by the JURI and PETI committees of the European Parliament. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/696968/IPOL_STU(2021)696968_EN.pdf

12 BBC. (2017). *Row over AI that "identifies gay faces"*. Available at: https://www.bbc.com/news/technology-41188560

13 BBC. (2020). *Facial recognition to "predict criminals" sparks row over AI bias*. Available at: https://www.bbc.com/news/technology-53165286

14 Barrett, L. F. et al. (2019). 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements'. *Psychological Science in the Public Interest*, 20(1). Available at: https://journals.sagepub.com/doi/10.1177/1529100619832930

these practices are currently being debated,[15] they have not been sufficiently proven. This makes uses of biometrics that claim to identify personality traits or emotional state currently no more robust than phrenological pseudo-science,[16] while putting the value of equality under strain and promoting questionable theories about individual identity.

Current evidence about the use of these technologies has shown their limitations, unreliability and the risk they pose to both individual fundamental rights and societal values. We strongly recommend that these uses of biometric technologies are added to the unacceptable and high-risk categories of use.

Action: We propose that the use of biometrics for categorisation and emotion recognition should be included within the scope of Article 5. If, in the future, evidence emerges to demonstrate the benefits of these technologies, they could be added to the list of technologies that can be used in exceptional circumstances, with appropriate justifications.

8. Make the deployment of any AI system that poses an 'unacceptable risk' possible only in exceptional circumstances where it passes a 'reinforced proportionality test'

We recognise the concerns that have been raised that the breadth of exceptions to prohibited AI systems could potentially lead to extensive use of biometric technologies. EDPB and EDPS, as well as numerous civil society organisations, have put forward persuasive arguments in favour of complete bans on any use of AI for automated recognition of human features in publicly accessible spaces, in any context.[17]

Should a ban not be adopted, we suggest implementing a clearer, more comprehensive system to ensure that the use of these systems is limited, thoroughly considered, authorised by a public body and proportionate.

In accordance with recommendations above, a clearer distinction should be drawn between the systems that are to be completely banned and those which can be used in exceptional circumstances.

---

15 Ali, M.R., Myers, T., Wagner E., et al. (2021). 'Facial Expressions can detect Parkinson's disease: preliminary evidence from videos collected online'. *Npj Digit. Med.*, 129(4). Available at: https://www.nature.com/articles/s41746-021-00502-8

16 Ajunwa I. (2022).'Automated Video Interviewing as the New Phrenology'. *36 Berkeley Tech*. L.J. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3889454

17 EDPB-EDPS. (2021). *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).* Available at: https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en; EDRi. (2020) 'Ban Biometric Mass Surveillance'. Available at: https://edri.org/our-work/blog-ban-biometric-mass-surveillance/

The current formulation of Article 5 makes it clear that some unacceptable AI practices are to be completely banned, for example those deploying subliminal techniques to cause physical or psychological harm (Article 5(1)(a)). Others could be permitted for use in exceptional circumstances, such as real-time remote biometric identification, which the Commission envisages can be used to search for missing children.

If an AI system that gives rise to unacceptable risks may be allowed to be both marketed and used, the threshold for any use should be raised: its deployment should be conditional on passing a reinforced proportionality test to prove that the benefits of its use outweigh the risks.

The principle of proportionality plays a key role in the EU legal framework. It is used by the Court of Justice of the European Union to determine the limits of fundamental rights, and the extent to which potential infringements to them could be justified. This test requires that any measure taken that might infringe a fundamental right has a legal basis (is provided by law), meets the objectives of its use, is necessary to achieve those objectives and does not go beyond what is needed to do so (including in extent and period of use).

The reinforced proportionality test should consider harms that are individual (for example discrimination against a person), collective (discrimination against minorities) and societal and environmental, and which could alter the fabric of society for example by affecting the way in which equality is perceived or threatens the rule of law.[18]

They should be used for all systems included within the scope of Article 5, where a case might be made for their use. This proposal extends the model in Article 5(3), which requires national judicial authorities to carry out a proportionality test before approving the use of remote real-time biometric identification by law-enforcement authorities.

When deployed, these exceptionally permitted systems must then be treated as high risk and meet all relevant obligations attached to this categorisation.

Action: Introduce, in Article 5 of the proposed Act, a 'reinforced proportionality' test, which goes beyond the risks to individual human rights, health and safety to also address societal and environmental harms, and to make any deployment of 'unacceptable risk' AI systems in exceptional circumstances conditional to passing the test.

18 Smuha, N. (2021). 'Beyond the individual: governing AI's societal harm'. *Internet Policy Review*, 10(3). Available at: https://policyreview.info/articles/analysis/beyond-individual-governing-ais-societal-harm

9. Include an obligation to publish all decisions that approve the placing on the market or deployment of unacceptable risk systems in exceptional circumstances

Public trust in AI is crucial for its uptake and success. Because the Act identifies certain forms of AI as posing unacceptable risks, it will be essential for the public to be able to scrutinise the justification for any specific instances of use to maintain trust in the regulation. This would require an obligation to publish all decisions about marketing and deploying prohibited AI systems, excepting any information thatis legitimately confidential under relevant domestic or European law.

Action: Make public every decision about the deployment or marketing of any AI system that is categorised as posing an unacceptable risk, to ensure transparency about these decisions and allow the public access to them.

10. Ensure Articles 5(1)(a) and 5(1)(b) offer meaningful and substantial protection, by removing the reference to 'subliminal', broadening the scope of 'vulnerability' beyond age and disability, and adding 'economic damage' to their scope

Article 5(1)(a) of the proposed Act currently defines subliminal techniques as those 'going beyond a person's consciousness', but this does not adequately clarify their scope, or separate subliminal techniques from manipulation. We propose that this article would be strengthened by removing the reference to subliminal, on the basis that manipulation, whether conscious or subconscious, has negative effects.

In Article 5(1)(b), the Commission takes a narrow view of vulnerability, categorising responsibilities only with regard to age and disability. While valid, these categories may require clarification and expansion, for example to acknowledge those with mental health issues, like depression, anxiety or addiction, and to consider whether other vulnerabilities should be added to the list.

Both paragraphs 5(1)(a) and 5(1)(b) should be extended to also apply to economic damage, to protect against techniques that drive people into gaming addictions, for example, with negative economic effects.

These Articles currently conflate safety with fundamental rights protection, which seems (as with other parts of the proposed Act) to be a legacy of the Commission's product-safety approach. As currently drafted, the two paragraphs consider safety requirements, but neither adequately considers fundamental rights protections, in particular dignity, which is protected

under Article 1 of the Charter of Fundamental Rights of the European Union,[19] or broader societal harms.

Action: Amend Article 5(1)(a) to remove 'subliminal' and Article 5(1)(b) to broaden the scope of vulnerability, and include economic damage in the scope of both.

Title III: High-risk AI systems

11. Create a mechanism to allow the Commission to add new categories of high-risk AI systems to Annex III

The proposed Act currently assumes that the list of categories of high-risk AI systems in Annex III is comprehensive and complete. It acknowledges the need to add new uses of technologies, but only allows the Commission to add new subcategories.

This brings partial futureproofing that is out of step with the nature of AI, which evolves quickly and significantly. A mechanism should, therefore, be put in place to ensure that the list of categories in Annex III can be extended, as well as retaining the ability to add subcategories below those existing categories.

The Commission should also enable the public to express their concerns and flag any systems that they believe should be added to this list. The Commission should have an obligation to consider these concerns, where relevant together with the EU AI Board, in a timely manner and present a reasoned response.

Action: Article 7(1) should be amended to reflect this change, in particular by removing the requirement in Article 7(1)(a).Add systemic and environmental risk to the list of criteria in Article 7

When assessing whether a system should be included in the high-risk category, the Commission must currently consider a number of criteria that are listed in Article 7, and cover risks to health, safety and fundamental rights. While valuable, as suggested in recommendations 6 and 10, this list of criteria must be expanded to include systemic and environmental risks.

19 Charter of Fundamental Rights of the European Union. OJ C 326, 26.10.2012, p. 391–407. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT

Action: Amend Article 7(2) to explicitly list systemic and environmental harms as elements the Commission must take into account when expanding the list of high-risk systems.

12. Add a requirement for all high-risk AI systems to be subjected to regular 'impact evaluations'

Chapter 2 of Title III of the proposed Act lists all the requirements that high-risk AI systems must meet. Currently these requirements are all considered *ex ante*, so are applicable before the system is put on the market. As above, this is the legacy of the product-safety approach taken by the Commission and does not reflect how AI systems behave in practice. We propose additional requirements are added *ex post*.

A facial recognition system that is tested for bias in lab settings, for example, may still be used in discriminatory ways once deployed and produce discriminatory results in response to new data. For example, a study done by Amnesty International shows how the use of facial recognition technologies reinforce systemic issues of over-policing of non-white neighbourhoods in New York.[20]

Even when the data used meets the standards laid out in Article 10 of the proposed Act, there is evidence that algorithms have still perpetuated biased and discriminatory outcomes. Examples range from 'algorithmic redlining' in financial services to the use of postcodes as proxies for race in assessing the risk of reoffending in the criminal justice system.[21]

We propose that repeated evaluations should be conducted by deployers using the system, in close cooperation with the providers, at regular intervals post-deployment. The appropriate intervals of time should be decided by the Commission, perhaps in consultation with the future EU AI Board. We recognise that this is a novel approach and are committed to continue working to develop evidence to support the best practical solutions for this proposal.

20 Amnesty International. (2022). 'USA: Facial recognition technology reinforcing racist stop-and-frisk policing in New York – new research'. Available at: https://www.amnesty.org/en/latest/news/2022/02/usa-facial-recognition-technology-reinforcing-racist-stop-and-frisk-policing-in-new-york-new-research/

21 Rovatsos, M., Mittelstadt, B., and Ansgar, K. (2019). 'Landscape Summary: Bias in Algorithmic Decision-Making'. *Centre for Data Ethics and Innovation*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf

Action: To address these situations, an ex post requirement should be added to the list. This would be an evaluation of the real-world impacts of a high-risk AI system, designed to ensure it is functioning as intended, that there are no errors or risks left unaddressed and that the system continues to meet the state-of-the-art standards required by the proposed Act.

13. **Add a requirement to ensure that high-risk AI systems are transparent for affected persons, in addition to the provision for an EU database for stand-alone high-risk systems in Article 60**

The proposed Act recognises a transparency obligation towards affected persons in only three instances, chatbots, deep fakes and emotion recognition, all provided for under Title IV. We propose that the transparency obligation towards affected persons should be extended to all high-risk systems. We are currently working on devising concrete ways in which this can be achieved in practice.

Action: This obligation, as expressed in Article 52(1), should be extended to all high-risk AI systems and not limited to the three systems contained therein.

Title IV: Transparency obligations for certain AI systems

14. **Remove Title IV completely as the systems it addresses fall either in the 'unacceptable' or 'high-risk' category, according to the list of criteria defined in recommendation 4**

This Title seeks to create a middle ground, for a number of systems that are seen as not being harmful enough to be included in a higher category of risk but nevertheless harmful enough to require specific obligations.

However, on the basis of the proposal for criteria discussed above, we anticipate the three systems currently contained in this title – emotion recognition, deep fakes and chatbots – would fall within the list of unacceptable or high-risk technologies, making this category redundant.

Action: Apply the transparency obligation in Article 52(1) to all high-risk systems, meaning that this Title will no longer be necessary. Where relevant, specific transparency obligations for each of these systems should be kept as part of the proposed Act but integrated within the appropriate risk category.

Titles VI, VII, and VIII: Governance and enforcement

15. **Create a comprehensive remedies framework for affected persons, including a right for individuals to bring complaints; a right to bring collective action; a right to information, supplementing what is already provided by the General Data Protection Regulation ('GDPR')**

In recommendation 1 we set out the consequences of affected persons not being included in the framework imagined by the proposed Act. The most significant omission is in the governance and enforcement of the future Act, neither of which meaningfully includes those ultimately affected.

There are currently no individual rights to information or to bring complaints, and no right to bring collective action. This means that the interests of the individuals who are ultimately affected by the use of AI systems are not built into the framework that is meant to regulate those systems.

Remedying this should be a priority for the European Parliament and for the Council of the European Union.

Action: Having included affected persons in the framework (see recommendation 1), they must also be granted the right to bring individual complaints before national Market Surveillance Authorities, and – where relevant – domestic courts, as well as a right to information supplementing what is already provided by the GDPR. Taking account of the nature of AI and its potential societal impacts, a right to bring collective action should also be included.

16. Build civil society representation into the mandated standard-setting process and place an obligation on the Commission to review these standards before they can be used

As the Commission itself has recognised, the standards set do not only relate to technical components but also 'incorporate core EU democratic values and interests'.[22]

The Commission is empowered by the proposed Act to mandate the development of harmonised standards for high-risk AI, most likely by CEN-CENELEC. Should that happen, the decision-making processes of these bodies should ensure that the standards are developed to also incorporate fundamental rights assessments and mitigation against systemic risks

Action: We propose that this process should be formally build in contributions from civil society organisations, as the Commission suggests it plans to do in the future23, and that the Commission has an obligation to review the harmonised standards for high-risk AI before adopting them.

17. Ensure the framework in place will provide Market Surveillance Authorities with the resources necessary to carry out their role and responsibilities

National Market Surveillance Authorities have a crucial role to play in enforcing the future Act. Once a right to bring individual complaints is added, their role will become even more important. To ensure that they can effectively consider and deal with such complaints, Market Surveillance Authorities must be provided with the relevant necessary resources, in particular to reflect their potential competence to deal with complaints, and encouraged to cooperate with other relevant authorities, such as data and consumer protection authorities.

Action: Amend the role of Market Surveillance Authorities to reflect the changes made to enforcement and clarify their role in dealing with individual complaints.

---

22 European Commission. (2022). *Communication - An EU Strategy on Standardisation - Setting global standards in support of a resilient, green and digital EU single market.* COM(2022) 31 final 2.2.2022. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022DC0031

23 European Commission. (2022). 'New approach to enable global leadership of EU standards promoting values and a resilient, green and digital Single Market'. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_661

Acknowledgements

This paper has benefited from the valuable work of civil society organisations, academics and other specialists in this area. The contributions we considered include responses to the European Commission's consultation on the EU AI Act Proposal, academic articles about the legal framework of the European Union and reports on the use of biometric technologies.

We are particularly grateful to our external reviewers, who read an early version of this piece and provided feedback, ideas and challenge to improve it further. Many thanks to Mathias Vermeulen (AWO); Sarah Chander and Ella Jakubowska (EDRi); Vanja Škorič and Francesca Fanucci (ECNL); Daniel Leufer (Access Now) and Risto Uuk (Future of Life Institute).

Regulating AI in Europe: four problems and four solutions

Read Professor Lilian Edwards' expert analysis on how to help create 'trustworthy AI', which balances proportionately the social interest in innovation and better delivery of public services from AI, with adverse impacts on fundamental rights and societal values – in line with the proposed AI Act: https://www.adalovelaceinstitute.org/report/regulating-ai-in-europe/

For more information on the Ada Lovelace Institute in Europe: https://www.adalovelaceinstitute.org/europe/

Ada Lovelace Institute
100 St John Street, London, WC1B 3JS
+44 (0) 20 7631 0566

Registered charity 206601

Website: adalovelaceinstitute.org
Twitter: @AdaLovelaceInst
Email: hello@adalovelaceinstitute.org