Expert explainer

Lilian Edwards

Professor of Law, Innovation and Society,
Newcastle University

The EU AI Act: a summary of its significance and scope

Ada Lovelace Institute

April 2022

Introduction

This explainer is intended to provide those interested in global AI regulation and unfamiliar with the EU AI Act with a description of its significance, scope and main points.

It will be particularly useful for UK and global policymakers, as the Act is likely to become, or at least aspires to be, a global standard – as the GDPR has become for personal data protection – and will of course be one model for any future UK post-EU bespoke legislation.

Timeline

The European Commission released the proposed Regulation on Artificial Intelligence (the EU AI Act) on 21 April 2021.[1]

Since then amendments have been proposed both by successive Council Presidencies and will be made by committees charged with reporting on various parts of the Act. There is no guarantee these will be accepted in the process, so they are only briefly referred to below where particularly significant.

On 11 April, the draft report by the Internal Market and Consumer Protection (IMCO) and Civil Liberties, Justice and Home Affairs (LIBE) committees will be presented to the European Parliament. This is an opportunity for MEPs to consider their approaches to potential amendments to the Act, with a deadline of 18 May.

On 11 July opinions will be received from the Legal Affairs (JURI) and Industry, Technology and Research (ITRE) committees, and in October each committee will vote on amendments.

[1] European Commission. (2021). *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. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

In November, there will be a plenary vote in the EU Parliament, and in December the Trilogues will begin.

The Council Presidency passes from France to the Czech Republic in July 2022, and the Act is likely to be passed into law during the Swedish Council Presidency in the first half of 2023.

International significance

9 key points:

1. The Act sets out harmonised rules for the development, placing on the market, and use of AI in the European Union (EU).

2. The Act draws heavily on the model of certification of 'safe' products used for many non-AI products in the New Legislative Framework (NLF) (see below) and is part of a set of EU draft proposals to regulate AI, including the Machinery Regulation and the announced but not yet released product liability reforms.

3. The Act needs to be read in the context of other major packages announced by the EU such as the Digital Services Act (DSA), the Digital Markets Act (DMA) and the Digital Governance Act (DGA). The first two primarily regulate very large commercial online platforms such as notably Google, Amazon, Facebook and Apple (GAFA).

4. These well-known tech giants although prominently fuelled by 'AI', are not the focus of the Act: instead, it is mainly, though not exclusively, aimed at public sector and law enforcement AI.[2]

5. The Act does not replace but will overlap with the protections offered by the General Data Protection Regulation (GDPR),[3] although the former's scope is more expansive and not restricted to personal data.

[2] Note that the Act does not affect the application of the (current) E-Commerce Directive or (future) DSA; Article 2(5).

[3] Hence already the issue of a report on interaction of the Act with the GDPR by European Data Protection Supervisor and European Data Protection. See: 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/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf.

6. Parts of the Act related to manipulation and deception also draw on the Unfair Commercial Practices Directive (UCPD).[4]

7. Existing consumer law and national laws, such as on tort, are also relevant.

8. Like the GDPR, the territorial jurisdiction of the Act is expansive: governing not just *users* of AI in the EU, but providers who place on the market or into service AI systems in the EU.[5] Thus, despite Brexit, the Act will be crucial to the UK AI industry's aspirations as exporters to the EU or providers of 'AI-as-a-Service'.[6]

9. The Act's extraterritorial reach is backed by big fines akin to the GDPR (maximum 6% global annual turnover).

For the UK, choices may be presented between the EU model of 'trustworthy AI' rooted in a tradition of strong consumer law protection, human rights and 'ethics',[7] versus competing notions from the US and China, the latter especially offering tempting outcomes for developers because of a lower bar for personal data collection and human rights protection.

China itself seems to be shifting, on paper at least, to a more regulated model for data and AI, although the aim may be more to protect the state from the power of its own tech platform sector (AliBaba, TenCent,

[4] European Commission. *Unfair Commercial Practices Directive*. Available at : https://ec.europa.eu/info/law/law-topic/consumers/unfair-commercial-practices-law/unfair-commercial-practices-directive_en.

[5] NB note special use of 'user' here, Article 3(4), discussed below.

[6] Cobbe, J. and Singh, J. (2021). 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges.' *Computer Law and Security Review*, volume 42. Available at: https://www.sciencedirect.com/science/article/pii/S0267364921000467.

[7] As well as the Act, Europe will adopt a Coordinated Plan to take a 'leading position in the development of human-centric, sustainable, secure, inclusive and trustworthy AI'. See: European Commission. (2021). Press Release: *Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682. Funding is allocated from the Digital Europe and Horizon Europe programmes, as well as the Recovery and Resilience Facility that foresees a 20% digital expenditure target, and Cohesion Policy programmes. The Act has been in the works since 2018 when preceded by *inter alia*, a Strategy, a White Paper and a report on 'the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics' which identified gaps in the product safety regime for AI. (See: European Commission. (2020). *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics.* Available at: https://ec.europa.eu/info/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics-0_en).

WeChat, etc.) than to protect individual rights. The US under the Biden administration may finally move towards stronger protection of privacy.

These political shifts may make it even more likely that the EU Act becomes an acceptable global model. Paradoxically, post-Brexit the UK may be drawn ideologically away from European approaches to regulation, especially in the wake of COVID-19, as we are seeing in current debates over the future of the UK GDPR.[8]

The Act will be overseen by a European AI Board (where the UK will not as a third country have representation) but the Board's role is advisory, and enforcement, if any (see below) will be primarily by national Market Surveillance Authorities.

Drawing on the legacy of the New Legislative Framework (NLF) for creating safe and secure consumer products,[9] there is a strong leaning in the Act towards governance by private self-regulation, and rule-making by technical standard-setting bodies operating outside of normal democratic processes that will be difficult for civil society and users to engage with.

[8] See: Taskforce on Innovation, Growth and Regulatory Reform (2021). *Independent report*. Available at: https://www.gov.uk/government/publications/taskforce-on-innovation-growth-and-regulatory-reform-independent-report; Department for Culture, Media & Sport. (2021). Data: a new direction. Available at: https://www.gov.uk/government/consultations/data-a-new-direction.

[9] European Commission. *New Legislative Framework*. Available at: https://ec.europa.eu/growth/single-market/goods/new-legislative-framework_en.

Scope

The *scope* of the Act is apparently very wide, covering systems developed with any of the approaches in Annex I (machine learning, logic and knowledge-based approaches, and statistical or Bayesian approaches) that can generate outputs such as content, predictions, recommendations, or decisions influencing 'environments they interact with' (Article 3(1) and Annex I).

This has caused some concern that the Act may be unfeasibly wide, extending to much of modern software.[10] Others have suggested that AI is a red herring and other significant types of software may be excluded. In reality, however, these debates are fairly academic as the operational impact of the Act is quite narrow, the main thrust relating to 'high-risk AI' which is quite closely delimited (below).

Veale and Borgesius suggest that the width of the ostensible scope of the Act (all AI systems, including limited and minimal risk) may severely restrict EU national competence to legislate for AI, even in areas of so-called minimal risk like private-sector targeted marketing, or limited risk like 'deepfakes', given likely interpretations of the Act as a maximum harmonisation.[11]

This is one problem that should not afflict the UK post-Brexit. However, it might be useful to have clarified that the Act is without prejudice to existing EU laws such as the GDPR (only the ECD/DSA is mentioned as unaffected in Article 2(5)).

Who is subject to obligations under the Act?

Predominantly, '*providers*' of systems who develop an AI system with a view to placing it on to the market or putting it into service under their own name or trademark (Article 3).

[10] AI systems developed exclusively for military purposes are however excluded.

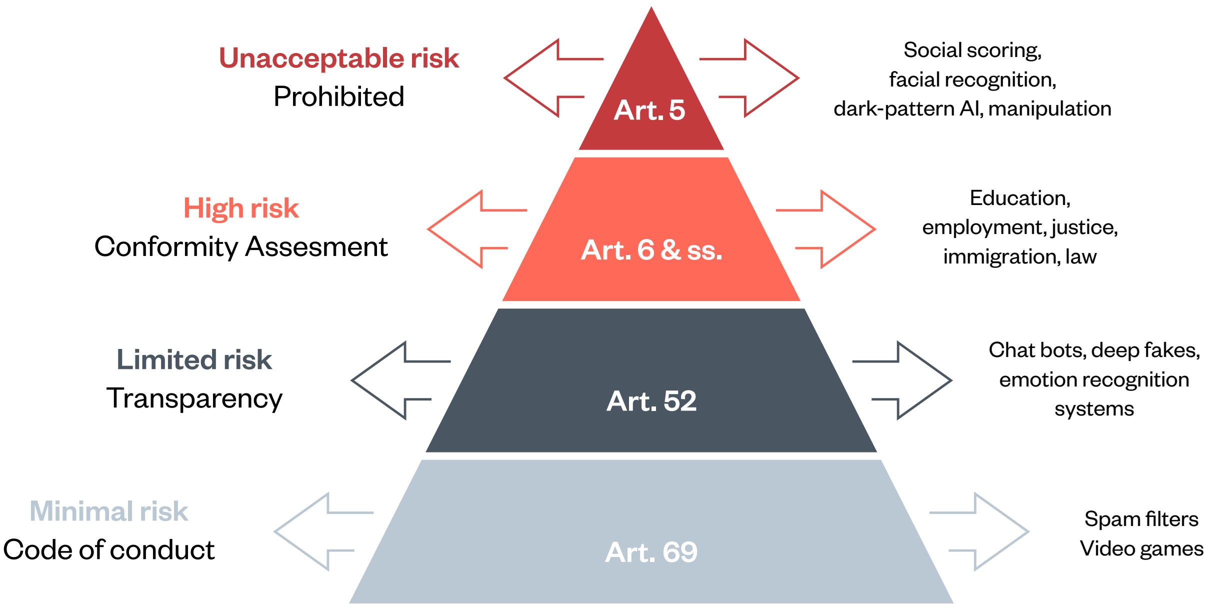
[11] Veale, M., and Borgesius, F. Z. (2021). 'Demystifying the Draft EU Artificial Intelligence Act'. *Computer Law Review International*, 22(4).

But obligations also, or alternatively, fall in different ways on ‘*users*’, defined here rather contra-intuitively to mean any natural or legal person ‘*using an AI system under its authorit*y’, e.g. a local authority putting in a fraud detection scheme, or an employer putting in an automated hiring system. This is *not* the ‘ultimate end user’ of the system, or the ‘data subject’ in the GDPR and there is no word in the Act for this person.

Obligations also fall on importers and distributors (Articles 26–28) in a way similar to the product safety regime, with the intent of stopping dangerous products built outside the EU from entering it. The primary actor on whom obligations are placed is nonetheless the provider.

While this scheme has worked relatively well for tangible products, where the manufacturer of a product is the analogue of the provider in the Act, this top-down allocation of duties will be more challenged in a world of ‘upstream’ AI services, which can be re-used downstream in a wide variety of unforeseeable contexts. We discuss this and ‘general purpose’ AI in particular, below.

Structure: a ‘risk-based’ approach



The Act splits AI into four different bands of risk based on the intended use of a system. Of these four categories, the Act is most concerned with ‘high-risk AI’.

Although described as a ‘risk based’ scheme, there is no sliding scale of risk, merely one category (‘high risk’), which is – at least on paper – extensively regulated; some minor transparency provisions for a small number of systems characterised as ‘limited-risk’ AI; and a number of ‘red lines’, which have rhetorical effect but in practice are likely to be of very limited application.

In descending levels of concern:

1. Unacceptable risk:

These are prohibited AI applications 'considered unacceptable as contravening Union values, for instance by violating fundamental rights'. These comprise:

- Subliminal techniques: AI 'that deploys *subliminal techniques... in order to materially distort a person's behaviour in a manner that causes or is likely to cause... physical or psychological harm*' (Article 5(1)(a)).

- Manipulation: 'an AI system that *exploits any 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 likely to cause... physical or psychological harm*';(Article 5(1)(b). Note: this is unlikely to cover 'dark patterns' in consent-management platforms or similar consumer online interactions, as there is highly unlikely to be '*intent*' to manipulate *in order to cause* physical or psychological harm.

- Social scoring: a social behaviour credit system created or used by a *public authority* to evaluate or classify the '*trustworthiness*' of natural persons; (Article 5(1)(c)). As this is restricted to the public sector it does not cover private-sector profiling and targeting. Even within the public sector AI it is not clear whether a system for assessing families at high risk for child neglect or abuse, for example, would fall within scope. Article 5(1)(c) is only invoked if data collected or generated for one purpose is 'misused' for profiling in a *different* social context; or alternately if the treatment is 'unjustified or disproportionate to their social behaviour or its gravity'.

- Biometrics: '*"real-time" remote biometric identification systems*' in '*publicly accessible spaces*' used by law enforcement [with major exceptions] (Articles 5(1)(d) and 5(2)(4)).

Biometrics

An area of controversy in the Act is whether it should include an effective total ban on use of facial recognition (or other types of biometric identification) and if so by what actors: law enforcement or private actors. At present under Article 5(1)(d) a ban exists but only in relation to:

- the targeted search for specific potential victims of crime, including missing children.
- the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack.
- 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/JHA62 and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State.'

This is a very limited ban. Only '*real-time*' biometric identification systems are banned, i.e., those that identify individuals at a distance by comparing the biometrics of the observed subject with a biometric database without 'significant delay' (Article 3(37)).

Publicly accessible spaces '*does not cover places that are private in nature and normally not freely accessible for third parties, including law enforcement authorities, such as homes, private clubs, offices, warehouses and factories*'. Online spaces are also not included.[12]

The restriction to law enforcement purposes excludes private security even though it may represent similar threats to fundamental rights. National security uses will also be excluded by virtue of Union law.[13]

[12] AI Act, recital 9.

[13] Forthcoming independent legal review on biometrics by Matthew Ryder QC.

The 'ban' imposed by the Act may sometimes be less stringent than existing data protection controls under GDPR and the Law Enforcement Directive (LED). Thus if the maximum harmonisation argument (above) operates, the Act might in fact reduce protection against biometric surveillance already given by existing national laws.

It is important to note the division drawn between *biometric ID systems* that uniquely identify individuals and *biometric categorisation systems* that put people into classes – male, young, Uyghur, headscarf-wearing, tattooed, etc.

Categorisation systems are only classed as 'limited risk' despite the heading of Annex III, containing the possibility that categorisation systems might in some future time be added to 'high risk'. The distinction between identity and categorisation seems to have been drawn for consistency with the definition of biometric data in GDPR Article 4(14), which requires that biometric data be data processed with intent to identify a data subject. However, the question here is not what the *purpose* was for which personal data was processed, but whether the risk of *impact* on fundamental rights is high – hence this distinction seems irrelevant, when the real issue is that biometric categorisation has been shown to create severe impact on the rights of surveilled groups.

Similarly, the processing of facial or other data (temperature, sweat, eye movements, etc.) to establish emotional states ('emotion recognition systems') is also only in principle regulated by Title IV ('limited risk'), though it may also fall into 'high risk' under Annex III, where used by law enforcement in limited circumstances (Annex II, 6(b)('polygraphs and similar tools').

This means private-sector emotion recognition systems are in principle not classified as 'high risk' by the Act, although uses in employment or education *might* be caught by Annex III (3) and (4).

A few points to note:

1. First, biometric ID systems, although not prohibited, explicitly remain 'high risk' when deployed in private-sector contexts, and/or not in real time (Annex III(1)).

2. They are the only category of 'high risk' AI system where prior authorisation must be given by a judicial or independent authority.

3. Under the 'human oversight' rules of Chapter 2 discussed below, (which apply to all high-risk systems) no action is to be taken on the basis of the identification given by a biometric system unless it is confirmed by two natural persons.

4. Member states must implement the biometric identification ban in national legislation, which can be more restrictive than the Act, either banning the use of such technology entirely or only allowing it for some of the three explicit situations.[14] This is thus the only case in the Act where maximum harmonisation rules will not prevent EU states from providing stronger protection to their citizens than other EU states may choose to do.

2. High risk

High-risk AI systems are subject to a detailed certification regime (see below), but are not deemed so fundamentally objectionable that they should be banned. These include:

1. In Annex II-A, AI systems intended to be used as a safety component of a product, or themselves a product, which are already regulated under the NLF (e.g. machinery, toys, medical devices) and, in Annex II-B, other categories of harmonised EU law (e.g. boats, rail, motor vehicles, aircraft, etc.).

2. In Annex III, an exhaustive list of eight 'new' high-risk AI systems, comprising:

[14] Forthcoming independent legal review on biometrics by Matthew Ryder QC.

a. **Critical infrastructures** (e.g. transport), that could put the life and health of citizens at risk

b. **Biometric ID systems** (see above)

c. **Educational and vocational training**, that may determine the access to education and professional course of someone's life (e.g. automated scoring of exams)

d. **Employment, workers management and access to self-employment** (e.g. automated hiring and CV triage software)

e. **Essential private and public services** (e.g. automated welfare benefit systems; private-sector credit scoring systems)

f. **Law enforcement systems that may interfere with people's fundamental rights** (e.g. automated risk scoring for bail; 'deepfake' law enforcement detection software; 'pre-crime' detection)

g. **Migration, asylum and border control management** (e.g. verification of authenticity of travel documents; visa processing)

h. **Administration of justice and democratic processes** (e.g. 'robo-justice'; automated sentencing assistance).[15]

3. The Commission can add new *sub-areas* to Annex III by delegated act if they pose an equivalent or greater risk than systems already covered, but cannot add entirely new top-level categories.

   The Slovenian Presidency compromise text suggested that insurance systems and critical infrastructure be added to 'high-risk' systems.

3. Limited risk

Three 'limited-risk' AI systems are exhaustively defined in Title IV (Article 52). The duties here are only to provide transparency such as labelling, or disclosure that content has been manipulated. The utility

[15] European Commission. (2021). Press Release: *Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682.

of this is debatable both technically, and in terms of overlap with GDPR, which already e.g. requires controllers processing personal data to be transparent about use of profiling and automated decision-making.

The systems are:

1. **Chatbots**
2. **Emotion recognition and biometric categorisation systems**
3. **Systems generating ‘deepfake’ or synthetic content**

In relation to *chatbots*, only providers, not users have, obligations of transparency. *Providers* must design a system such that users are informed that they are interacting with a machine rather than a human. If company A codes a chatbot app, it is only Company A, not the website B that has bought the chatbot software, which has to ensure transparency.

By contrast, in relation to emotion ID and deepfakes, duties only fall on *users* to provide transparency. Interestingly, even well before the Act has passed, Twitter has introduced a voluntary scheme where operators of tweetbots can mark themselves as automated.[16]

4. Minimal risk

Minimal risk includes applications such as spam filters or AI-enabled video games. The Commission proposes that these are mainly regulated by voluntary codes of conduct.[17]

[16] See: Perez, S. (2021). ‘Twitter introduces a new label that allows the “good bots” to identify themselves’. *TechCrunch*. Available at: https://techcrunch.com/2021/09/09/twitter-introduces-a-new-label-that-allows-the-good-bots-to-identify-themselves.

[17] Explanatory memorandum 5.2.7; Article 69.

‘Essential requirements’ for high-risk AI

The Act requires providers of high-risk AI systems to conduct a prior conformity assessment before placing them on to the market (Articles 16 and 43). Providers, in line with the NLF model, must ensure their systems are compliant with the **‘essential requirements’** set out in Title III, Chapter 2 of the Act. They can then attach a CE mark to conforming systems, which can be freely imported and distributed throughout the EU.

These requirements relate to **data and data governance; technical documentation; record keeping; transparency and provision of information to users; human oversight; and robustness, accuracy and security**.

Providers must set up a risk-management system that documents and manages risks across the AI system’s entire lifecycle, when used as intended, or, under conditions of ‘reasonably foreseeable misuse’.[18] Risks may be added as a result of post-market surveillance (see below). The aim is to bring down the ‘high risks’ of the AI system to an acceptable residual level. ‘Adequate mitigation and control measures’ can be used when risks cannot be eliminated. Residual risks are to be communicated to users.

Two categories of requirements seem particularly germane to the principal worries in the literature around AI in our society making decisions that affect humans, namely: algorithmic error, bias and discrimination; automated decision-making as contrary to human dignity; and opacity/lack of explanations.

[18] Article 9.

1. Data and data governance (Article 13)

This aims to meet a set of concerns about the quality of the data used to build AI systems and includes:

- Rules about how *training sets* (also validation and testing datasets) must be designed and used. Significantly for concerns about error and discrimination generated by partial, erroneous or historically biased data, datasets must be *‘relevant, representative, free of errors and complete’*, taking into account the intended purpose.[19]

- Rules about data preparation, including ‘annotation, labelling, cleaning, enrichment and aggregation’.

- Assessment of the ‘formulation of relevant assumptions [about] the information that the data are supposed to measure and represent’; and ‘examination in view of possible biases’.

- Exemption from GDPR rules restricting collection of sensitive personal data to de-bias algorithms.

2. Human oversight (Article 14)

Systems must be designed and developed in such a way that they can be *‘effectively overseen by natural persons during the period in which the AI system is in use’*.[20]

This is not simply a matter of transparency or explanation of how the AI system ‘works’, as discussed in the context of Article 22 and 13–15 of the GDPR, but goes much further into terrain such as enabling the ‘human overseer’ to spot anomalies, become aware of ‘automation bias’, be able to correctly interpret the system’s outputs and be able to override or disregard the system. Explicitly, one aim is to prevent or minimise risks to fundamental rights (Article 14(2)).

[19] Article 10(3). There has been much pushback from industry on the idea that datasets could ever be ‘error free’ and it is likely the final text will be more moderate.

[20] Article 14(1).

If a high-risk system is operated by a 'user' rather than the original provider – e.g. a private company buys and installs an automated hiring system from HireVue – then the allocation of responsibility is quite different in the Act than in the GDPR.

In GDPR language, the company would be the 'data controller' and the main focus of duties. In the Act, it remains the sole responsibility of HireVue to obtain conformity assessment before the system is put on the market,[21] and to implement 'human oversight' in a way that is 'appropriate to be implemented by the user'.[22] Under Article 28 if the user substantially modifies the system, they become the new 'provider' with all corollary certification duties.

General-purpose AI

As demonstrated above, the Act primarily puts obligations on upstream providers as opposed to downstream users, unless those users make a 'substantial modification' and become treated as new providers. This creates problems when talking about what has been labelled *general-purpose AI*. This is not 'artificial general intelligence' (AGI) – AI that has human level sentience and autonomy– but AI that has multiple possible uses in different contexts. Key examples include large language models used to generate text or speech, recognise text or speech, analytics, language translation modules etc. These are often included in downstream applications via APIs or 'calls' – 'AI as a service'.[23]

The Act's classification of systems as high-risk is based on *intended use* in one of the listed sub-areas of categories (see Annex III); so general-purpose AI does not fit in well. As such it is possible that no obligations to comply with Chapter 2's essential requirements fall on their providers, which seems unfortunate, given these systems may generate unfair and discriminatory outcomes.

[21] Article 14(3)(b).

[22] Article 14(3)(a).

[23] Cobbe, J. and Singh, J. (2021). 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges.' *Computer Law and Security Review*, volume 42. Available at: https://www.sciencedirect.com/science/article/pii/S0267364921000467.

The question then arises if a downstream user deploying a general-purpose AI system as intended – e.g. to generate speech-from-text – makes a 'substantial modification'? (Article 3(23)). If they do not, they are not liable as a provider either, which means these systems are completely unregulated by the Act.

The Slovenian Presidency added a compromise text on this matter. A new Article 52A was suggested,[24] concerning 'general-purpose AI',[25] which seems to remove any 'high-risk' AI obligations in relation to 'general-purpose' AI from upstream *providers*. By contrast a *user* who '*integrates a general-purpose AI system made available on the market, with or without modifying it*' is deemed to become the provider and thus may be liable for certifying the system meets the Act's requirements. Some regard this as unsatisfactory given the provider, not the user, is most likely the actor with the power to review and modify the system's design, including altering its training or testing datasets. This is a provision quite likely to morph as the Act progresses.

[24] See: Council of the European Union. Note: *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 - Presidency compromise text*. 2021/0106(COD). Available at: https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf.

[25] Defined in new recital 70a as 'AI systems that are able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation etc.'.

Conformity assessment (pre-marketing) and enforcement (post-marketing)

Pre-market

The Act requires providers to ensure before placing on market that their systems conform with the essential requirements listed above, as well as to comply with a number of other tasks including registering AI systems on a database, having an appropriate quality management system in place,[26] recording the system's technical documentation and keeping automatically generated logs. After doing so, their system gets its CE mark, which enables distribution throughout the EU (Article 19).

Providers in the main will only have to demonstrate conformity by an 'assessment based on *internal control*' i.e. **self-certification** (Article 43(1)(a). All providers need to do is self-assess that their quality management system, technical documentation, and post-market monitoring plan follow the essential requirements. They can do this either by their own bespoke plans for conformity, or, much more likely, by following a relevant harmonised technical standard (see below).[27] At present, only a subset of high-risk AI systems must make use of a third-party body – a 'notified body' – to externally audit their conformity.

The systems in question are:

1. AI systems for biometric identification or categorisation of natural persons (Article 43(1)) *but only if* no technical harmonised standard is made, which is unlikely to result.

[26] 'Appropriate' to mean complying with Article 17.

[27] For more detail, see: Veale, M., and Borgesius, F. Z. (2021). 'Demystifying the Draft EU Artificial Intelligence Act'. *Computer Law Review International*, 22(4), pp.15-16.

2. AI systems already regulated under existing NLF or other EU laws, listed in Annex II, where that legislation already demands a notified body be involved (Article 43(3)).

How do providers self-certify? The Act anticipates that harmonised technical standards for high-risk AI will be created by technical committees. High-risk AI systems which self-certify as confirming with such standards are then *presumed* to have met the requirements of Chapter 2 (essential requirements)(see Article 40). Providers can ignore these standards, and instead justify that they have adopted technical solutions at least equivalent – but why would they, compared to the ease of using the official standard, which has already done the work of 'translating' the essential requirements into clear instructions?

A key issue then will be whether these technical standards will truly embody the substantive goals of the Act, especially regarding fundamental rights. Standards will be created for the EU by the standards bodies CEN and CENELEC. CEN/CENELEC have established an AI Technical Committee (TC) which expects ANEC to represent the views of consumers and who are active in digital rights,[28] and also has representation for workers and the environment.

An issue for non-EU states is that they are likely to have no voice or at least no vote on these EU standards,[29] but will find themselves subject to them when they export to the EU. An even more profound issue is that the EU standards may by reciprocity or osmosis become global standards (e,g, via ISO). So just as the Act may become a global regulatory model, these standards may also become global rules in effect.

[28] See: https://www.cencenelec.eu/. Info. from private email from Laurens Hernalsteen, rapporteur to the TC, on 19 July 2021.

[29] See potential reforms in the EU Standardisation Strategy: European Commission. (2022). Press Release: *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.

Post-market

What about enforcement after the system has gone on to the market? Who if anyone checks that the system is being used correctly, or if it has changed and learned, or been modified, that it is still in conformity with the essential requirements?

Providers are tasked to 'establish and document a post-market monitoring system in a manner that is proportionate to the nature of the artificial intelligence technologies and the risks of the high-risk AI system'.[30] This monitoring system will 'collect, document and analyse relevant data provided by users or collected... throughout their lifetime'.[31] Users (i.e. deployers) are also tasked to monitor systems 'on the basis of the instructions of use' and report new risks, serious incident or 'malfunctioning' (Article 29(4)). Their intel goes to providers or distributors *not* directly to the Market Surveillance Authority (MSA) (see below).

Provider monitoring of accidents and malfunctions must go to the relevant **Market Surveillance Authority (MSA)** at least within 15 days of becoming aware of it (Article 62). Member states are to appoint national supervisory authorities, which by default act as MSAs (Article 59) though in some cases other bodies such as Data Protection Authorities (DPAs) will take the role.

These are, unlike notified bodies, public bodies with regulatory power e.g. to require access to training, validation and testing datasets used by the provider, and the AI source code.[32] They have considerable power including to withdraw products and oblige intermediaries to cooperate in withdrawing the products from the market.[33] If risks to fundamental rights are present, the MSA is also to inform the relevant national public body or regulator (e.g. the state DPA) (Article 65). If MSAs either lack sufficient expertise or resources to act, or are unwilling to do so, the Commission itself can intervene to secure a consistent application of the law. This is a very residual power though and it should be noted that the EU AI Board as currently conceived is in no way a central EU 'super regulator'.

[30] Article 61(1).

[31] Article 61(2).

[32] Article 64(2).

[33] See: Market Surveillance Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC, and Regulations (EC) No. 765/2008 and (EU) No. 305/2011.

For more information about the EU AI Act, see ‘Regulating AI in Europe.’[34]

For 18 recommendations to strengthen the EU AI Act, see ‘People, risk and the unique requirements of AI’.[35]

[34] 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/

[35] Ada Lovelace Institute. (2022). Policy briefing: *People, risk and the unique requirements of AI*. Available at: https://www.adalovelaceinstitute.org/policy-briefing/eu-ai-act/

About the author

Lilian Edwards is a leading academic in the field of Internet law. She has taught information technology law, e-commerce law, privacy law and Internet law at undergraduate and postgraduate level since 1996 and been involved with law and artificial intelligence (AI) since 1985.

She worked at the University of Strathclyde from 1986–1988 and the University of Edinburgh from 1989 to 2006. She became Chair of Internet Law at the University of Southampton from 2006–2008, and then Professor of Internet Law at the University of Sheffield until late 2010, when she returned to Scotland to become Professor of E-Governance at the University of Strathclyde, while retaining close links with the renamed SCRIPT (AHRC Centre) at the University of Edinburgh. She resigned from that role in 2018 to take up a new Chair in Law, Innovation and Society at Newcastle University.

She is the editor and major author of *Law, Policy and the Internet*, one of the leading textbooks in the field of Internet law. She won the Future of Privacy Forum award in 2019 for best paper ('Slave to the Algorithm' with Michael Veale) and the award for best non-technical paper at FAccT in 2020, on automated hiring. In 2004 she won the Barbara Wellberry Memorial Prize in 2004 for work on online privacy where she invented the notion of data trusts, a concept which ten years later has been proposed in EU legislation. She is a partner in the Horizon Digital Economy Hub at Nottingham, the lead for the Alan Turing Institute on Law and AI, a Turing fellow, and a fellow of the Institute for the Future of Work. At Newcastle, she is the theme lead in the data NUCore for the Regulation of Data. Edwards has consulted for, inter alia, the EU Commission, the OECD, and WIPO. In 2021-22, she is part-seconded to the Ada Lovelace Institute to lead their work on the future of global AI regulation.

About the Ada Lovelace Institute

The Ada Lovelace Institute was established by the Nuffield Foundation in early 2018, in collaboration with the Alan Turing Institute, the Royal Society, the British Academy, the Royal Statistical Society, the Wellcome Trust, Luminate, techUK and the Nuffield Council on Bioethics.

The mission of the Ada Lovelace Institute is to ensure that data and AI work for people and society. We believe that a world where data and AI work for people and society is a world in which the opportunities, benefits and privileges generated by data and AI are justly and equitably distributed and experienced.

We recognise the power asymmetries that exist in ethical and legal debates around the development of data-driven technologies, and will represent people in those conversations. We focus not on the types of technologies we want to build, but on the types of societies we want to build.

Through research, policy and practice, we aim to ensure that the transformative power of data and AI is used and harnessed in ways that maximise social wellbeing and put technology at the service of humanity.

We are funded by the Nuffield Foundation, an independent charitable trust with a mission to advance social well-being. The Foundation funds research that informs social policy, primarily in education, welfare and justice. It also provides opportunities for young people to develop skills and confidence in STEM and research. In addition to the Ada Lovelace Institute, the Foundation is also the founder and co-funder of the Nuffield Council on Bioethics and the Nuffield Family Justice Observatory.

Find out more:

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

Preferred citation: Lilian Edwards. (2022).
The EU AI Act proposal. Ada Lovelace Institute.
Available at: https://www.adalovelaceinstitute.org/resource/eu-ai-act-explainer/