

# Demystifying the Draft EU Artificial Intelligence Act

Michael Veale

*Faculty of Laws, University College London, United Kingdom*

Frederik Zuiderveen Borgesius

*Interdisciplinary Hub for Security, Privacy and Data Governance, Radboud University, The Netherlands*

Pre-print, July 2021. Version 1.1.

*Abstract* In April 2021, the European Commission proposed a Regulation on Artificial Intelligence, known as the AI Act. We present an overview of the Act and analyse its implications, drawing on scholarship ranging from the study of contemporary AI practices to the structure of EU product safety regimes over the last four decades. Aspects of the AI Act, such as different rules for different risk-levels of AI, make sense. But we also find that some provisions of the draft AI Act have surprising legal implications, whilst others may be largely ineffective at achieving their stated goals. Several overarching aspects, including the enforcement regime and the effect of maximum harmonisation on the space for AI policy more generally, engender significant concern. These issues should be addressed as a priority in the legislative process.

1. Introduction.....	2
<b>Context</b> .....	2
<b>Structure and Approach</b> .....	3
2. Title II: Unacceptable risks.....	3
<b>Manipulative systems</b> .....	3
<b>Social scoring</b> .....	6
<b>Biometric systems</b> .....	7
3. Title III Regime: High-Risk Systems .....	9
<b>Scope</b> .....	9
<b>The AI Act in the context of the New Legislative Framework (NLF)</b> .....	10
<b>Essential requirements and obligations</b> .....	10
<b>Conformity Assessment</b> .....	13
<i>Harmonised standards &amp; European Standardisation Organisations</i>	13
<i>Controversies of harmonised standards</i>	14
<i>Self-assessment and the (limited) role of Notified Bodies</i>	15
4. Title IV: Specific Transparency Obligations.....	16
<b>'Bot' disclosure</b> .....	16
<b>Emotion recognition and biometric categorisation disclosure</b> .....	17
<b>Synthetic content ('deep fake') disclosure</b> .....	18
5. Harmonisation and Pre-Emption.....	19
<i>Marketing</i>	20
<i>Use</i>	21
6. Post-marketing controls and enforcement .....	22
<i>Notification Obligations and Complaints</i>	23
<i>Database of Standalone High-Risk AI Systems</i>	24
7. Concluding Remarks.....	25

Thanks to Valerio De Stefano, Reuben Binns, Jeremias Adams-Prassl, Barend van Leeuwen, Aislinn Kelly-Lyth, Lilian Edwards, Natali Helberger, Christopher Marsden, Sarah Chander for comments and/or discussion; and the conveners and participants of several workshops including one convened by Margot Kaminski, one by Burkhard Schäfer, one part of the 2nd ELLIS Workshop in Human-Centric Machine Learning; and one between Oxford, KU Leuven and UCL.

# 1. Introduction

On 21 April 2021, the European Commission presented a proposal for a Regulation concerning artificial intelligence (AI), — the AI Act, for short.<sup>1</sup> The AI Act seeks to lay down harmonised rules for the development, placement on the market and use of AI systems which vary by characteristic and risk, including prohibitions and a conformity assessment system adapted from EU product safety law.

In this paper, we analyse the initial Commission proposal — the first stage in a potentially long law-making process.<sup>2</sup> The AI Act is sufficiently complex to prevent us from summarising it exhaustively. We instead aim to contextualise and critique it, and increase accessibility of the debate to stakeholders who may struggle to apply their expertise and experience to what at times can be an arcane proposal.

## Context

The first public indication of regulatory action of the type proposed in the AI Act were a cryptic few sentences found in the previous European Commission’s contribution to the Sibiu EU27 leader’s meeting in 2019.<sup>3</sup> Subsequently, then-President-Elect von der Leyen’s political guidelines for the Commission indicated an intention to ‘put forward legislation for a coordinated European approach on the human and ethical implications of Artificial Intelligence’<sup>4</sup> — the spark that the draft Act acknowledges as its genesis.<sup>5</sup> The proposed Regulation is part of a tranche of proposals which must be understood in tandem, including:

- the draft Digital Services Act (with provisions on recommenders and research data access);<sup>6</sup>
- the draft Digital Markets Act (with provisions on AI-relevant hardware, operating systems and software distribution);<sup>7</sup>
- the draft Machinery Regulation<sup>8</sup> (revising the Machinery Directive in relation to AI, health and safety, and machinery);
- announced product liability revision relating to AI;<sup>9</sup>
- the draft Data Governance Act (concerning data sharing frameworks).<sup>10</sup>

<sup>1</sup> 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) (hereafter ‘AI Act’). We will refer to this version as the ‘AI Act’ in footnotes brevity even though it is in draft.

<sup>2</sup> For readers unfamiliar with the European legislative process: the Commission is the European Union’s executive, and has a monopoly on policy initiative. Drafts are amended and adopted through a bicameral procedure between the directly elected European Parliament and the Council, which represents Member State governments. This procedure encompasses both formal stages and informal back-room compromise (trialogue).

<sup>3</sup> European Commission, ‘Europe in May 2019: Preparing for a More United, Stronger and More Democratic Union in an Increasingly Uncertain World’ (Contribution to the informal EU27 leaders’ meeting in Sibiu (Romania) on 9 May 2019, 9 May 2019) 33.

<sup>4</sup> Ursula von der Leyen, ‘A Union that Strives for More: My Agenda for Europe’ (Political Guidelines for the Next European Commission 2019-2024, 2019).

<sup>5</sup> AI Act, p.l.

<sup>6</sup> European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM(2020) 825 final).

<sup>7</sup> European Commission, Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) (COM(2020) 842 final).

<sup>8</sup> European Commission, Proposal for a Regulation of the European Parliament and of the Council on machinery products (COM(2021) 202 final) (Machinery Regulation).

<sup>9</sup> See European Commission, ‘Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Fostering a European Approach to Artificial Intelligence (COM(3032) 205 Final)’ (21 April 2021) 2.

<sup>10</sup> European Commission, Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) (COM(2020) 767 final).

## Structure and Approach

The ‘Act’ is a regulation based on Article 114 of the Treaty on the Functioning of the European Union (TFEU), which concerns the approximation of laws to improve the functioning of the internal market. The proposal mixes reduction of trade barriers with broad fundamental rights concerns in a structure unfamiliar to many information lawyers. While it may look new, much of the Act’s wording is drawn from a 2008 Decision establishing a framework for certain regulations concerning product safety, used in a wide array of subsequent legislation.<sup>11</sup> The main enforcement bodies of the AI Act, ‘market surveillance authorities’ (MSAs), are also common in EU product law. All this brings a range of novelties and tensions we will explore.

The Commission distinguishes different risk levels regarding AI practices, which we adapt to analyse in four categories: i) unacceptable risks (Title II); ii) high risks (Title III); iii) limited risks (Title IV); iv) minimal risks (Title IX). We cover each in turn, except for minimal risks, where Member States and the Commission merely ‘encourage’ and ‘facilitate’ voluntary codes of conduct.<sup>12</sup> We finally look at broader themes raised by the Act, in particular the important question of pre-emption and residual competences of Member States, and enforcement.

## 2. Title II: Unacceptable risks

Unacceptable risks attract outright or qualified prohibitions in the Act. Whether the AI Act would contain prohibited practices has been a matter of controversy. In 2018, the Commission set up a ‘High-Level Expert Group on AI’ to advise on its AI strategy. Members soon described industry pressure that led to the group dropping terms including ‘red lines’ and ‘non-negotiable’ from their policy recommendations.<sup>13</sup> A leaked version of the Commission *White Paper on Artificial Intelligence* contained a moratorium on facial recognition, controversially later expunged from the final version.<sup>14</sup>

The Commission’s proposal contains four prohibited categories, three prohibited in their entirety (two on manipulation, one on social scoring); and the last, ‘real-time’ and ‘remote’ biometric identification systems prohibited except for specific law enforcement purposes if accompanied by an independent authorisation regime.

### Manipulative systems

Two prohibited practices claim to regulate manipulation.<sup>15</sup>

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

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

<sup>12</sup> AI Act, art 69.

<sup>13</sup> Thomas Metzinger, ‘Ethics Washing Made in Europe’, *Der Tagesspiegel* (18 April 2019) <<https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html>> accessed 15 August 2019. Two-thirds of HLEG-AI members were industry representatives. See generally Michael Veale, ‘A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence’ [2020] *European Journal of Risk Regulation*.

<sup>14</sup> Access Now, ‘Europe’s Approach to Artificial Intelligence: How AI Strategy is Evolving’ (December 2020) 24–25 <<https://perma.cc/X3JM-2M6A>>.

<sup>15</sup> AI Act, arts 5(1)(a-b); described as manipulation in recital 15; pages 12–13.

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;

In briefings on the prohibitions, the Commission has presented an example for each. They border on the fantastical. A cross-over episode of *Black Mirror* and the Working Time Directive exemplifies the first: '[a]n inaudible sound [played] in truck drivers' cabins to push them to drive longer than healthy and safe [where] AI is used to find the frequency maximising this effect on drivers'. The second is a '[a] doll with integrated voice assistant [which] encourages a minor to engage in progressively dangerous behavior or challenges in the guise of a fun or cool game'.<sup>16</sup>

These provisions jar with a common understanding of manipulation. Manipulation can be understood through four necessary, cumulative conditions: the manipulator wants to *intentionally* but *covertly* make use of another's decision-making to *further their own ends* through exploiting some *vulnerability* (understood broadly).<sup>17</sup> The Act's provisions echo some of these conditions. The Act requires *intent* ('in order to'). It is limited to certain *vulnerabilities*, either caused by 'age, physical and mental disability' or exposed through 'subliminal techniques'. If reliant on subliminal techniques, they must be *covert* ('beyond a person's consciousness'). However, a final trigger is not whether a would-be manipulator's own ends are furthered, but instead on whether the activity 'causes or is likely to cause that person or another person physical or psychological harm'. This heavily limits the provision's scope.

Manipulative AI systems appear permitted insofar as they are unlikely to cause an individual (not a collective) 'harm'. This harm requirement entails a range of problematic loopholes. A cynic might feel the Commission is more interested in prohibitions' rhetorical value than practical effect.

In real life, harm can accumulate without a single event tripping a threshold of seriousness, leaving it difficult to prove.<sup>18</sup> These 'cumulative' harms are reinforced over time by their impact on individuals' environments, with hyperpersonalisation, engagement and 'dwell' metrics and impact on children often called out in this regard.<sup>19</sup> Indeed, manipulation in other fields of law leaves the AI Act already looking dated. Law in intimate partner violence increasingly considers underlying dynamics rather than one-off events.<sup>20</sup> Moreover, the AI Act explicitly excludes systems where distortion or harm arises from dynamics of the user-base entwined with an AI system,<sup>21</sup> excluding salient areas such as discriminatory ratings or recommendations on dating apps and online markets.<sup>22</sup>

<sup>16</sup> See, from DG CONNECT, Gabriele Mazzini, 'A European Strategy for Artificial Intelligence' (*2nd ELLIS Workshop in Human-Centric Machine Learning (YouTube recording)*, 10 May 2021) <<https://youtu.be/OZtuVKWqh10?t=10346>> accessed 22 June 2021, at 2:52:26 *et seq.*

<sup>17</sup> Marijn Sax, 'Between Empowerment and Manipulation: The Ethics and Regulation of For-Profit Health Apps' (PhD Thesis, Universiteit van Amsterdam (UvA) 2021) 110–12.

<sup>18</sup> See e.g. Oscar H Gandy, *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage* (Routledge 2009).

<sup>19</sup> See generally Nick Seaver, 'Captivating Algorithms: Recommender Systems as Traps' (2019) 24 *Journal of Material Culture* 421. Harms are identified especially in relation to children, see e.g. Beeban Kidron and others, 'The Cost of Persuasive Design' (*5 Rights Foundation*, June 2018) <<https://5rightsfoundation.com/uploads/5rights-disrupted-childhood-digital-version.pdf>>.

<sup>20</sup> See generally Evan Stark and Marianne Hester, 'Coercive Control: Update and Review' (2019) 25 *Violence Against Women* 81 (on how the concept of *coercive control* entered English law due to how 'discrete, injurious assaults [were] too narrow to capture [patterns] of coercion').

<sup>21</sup> AI Act, recital 16 ('intention may not be presumed if the distortion of human behaviour results from factors external to the AI system which are outside of the control of the provider or the user').

<sup>22</sup> Jevan A Hutson and others, 'Debiasing Desire: Addressing Bias & Discrimination on Intimate Platforms' (2018) 2 *Proc ACM Hum-Comput Interact* 73:1; Karen Levy and Solon Barocas, 'Designing against Discrimination in Online Markets' (2017) 32 *Berkeley Tech LJ* 1183.

Furthermore, an AI system may classify people (e.g. emotionally), while a *separate* downstream actor uses that classification harmfully.<sup>23</sup> How and to whom should the Act's prohibition apply? Upstream classification with both useful and harmful potency is a difficult-to-govern 'dual use' artefact familiar in technology policy.<sup>24</sup> Yet digital intermediaries frequently benefit from a mix of illegal and legal activity, for example in advertising or copyright.<sup>25</sup> The AI Act does not rise to this challenge.

Even where these prohibitions apply, they add little to existing EU law. Both resemble the Unfair Commercial Practices Directive, which prohibits commercial practices if they 'materially [distort] or [are] likely to materially distort the economic behaviour with regard to the product of the average consumer [...] or of the average member of the group'.<sup>26</sup> In that Directive, the latter condition is triggered if a vulnerability on the basis of 'physical infirmity, age or credulity' is foreseeable.<sup>27</sup> Commercial practices are broad, including advertising, communications and other 'acts' relating to goods or services.

The importance of the AI Act's expansion beyond the Unfair Commercial Practices Directive to *non-economic* decision-making<sup>28</sup> is limited by the Act's harm requirements. Legislators may wish to note that workable alternatives exist to harm tests in information law, such as 'reasonable person' requirements creating flexible red lines, which may strike a fairer balance in these complex situations.<sup>29</sup>

Lastly, unlike the Unfair Commercial Practices Directive which focusses on use, the AI Act also prohibits the *sale* of in-scope manipulative systems, e.g. to oppressive regimes. Yet vendors can attempt to dodge this requirement by selling general purpose AI systems which can be (re)configured by a user. The recitals indicate that the manipulation provisions relate to systems '*intended to distort human behaviour*' [emphasis added].<sup>30</sup> Few vendors would admit to such intention. Disguising the 'true' market for digital products is already common practice – consider stalkerware disguised as child trackers.<sup>31</sup> Reconfiguration (or to use the industry term, 'democratisation') of AI is a significant trend, typified by *AI-as-a-service*.<sup>32</sup>

<sup>23</sup> See, on multi-stage profiling, Reuben Binns and Michael Veale, 'Is That Your Final Decision? Multi-Stage Profiling, Selective Effects and Article 22 of the GDPR' (Presented at PLSC-EU 2019, on file with authors 2021). Classification itself may be considered a form of harm (of representation), but there is little legal protection around this. See Reuben Binns, 'Fairness in Machine Learning: Lessons from Political Philosophy' (2018) Conference on Fairness, Accountability and Transparency (FAT\* 2018), 8.

<sup>24</sup> See John Forge, 'A Note on the Definition of "Dual Use"' (2010) 16 *Sci Eng Ethics* 111.

<sup>25</sup> See e.g. coordinating intermediaries in real-time bidding, Michael Veale and Frederik Zuiderveen Borgesius, 'Adtech and Real-Time Bidding under European Data Protection Law' [2021] *German Law Journal*; Cristiana Santos and others, 'Consent Management Platforms Under the GDPR: Processors and/or Controllers?' in *Privacy Technologies and Policy* (Cham, Nils Gruschka and others eds, Springer International Publishing 2021).

<sup>26</sup> Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), OJ L 149/22, art 5.

<sup>27</sup> *ibid*, art 5.

<sup>28</sup> The limits of the UCPD to transactional decisions relate to its the desire to have both a harmonised yet open-ended definition of fairness, which would not be a mechanism for the Member States with a history of moral standards in consumer law to reintroduce them through creative interpretation and create barriers to European trade. See generally Hans-Wolfgang Micklitz, 'Unfair Commercial Practices and Misleading Advertising' in Hans-Wolfgang Micklitz and others (eds), *Understanding EU Consumer Law* (Intersentia 2009).

<sup>29</sup> For example, in Canada the federal private sector privacy law PIPEDA utilises flexible red lines. It stipulates that an 'organization may collect, use or disclose personal information only for purposes that a reasonable person would consider are appropriate in the circumstances' – a provision the regulator describes as establishing 'no-go zones'. See Personal Information Protection and Electronic Documents Act 2000 (Canada) s 5(3).

<sup>30</sup> AI Act, recital 16. Intention gets three mentions in this recital alone.

<sup>31</sup> Diarmaid Harkin and others, 'The Commodification of Mobile Phone Surveillance: An Analysis of the Consumer Spyware Industry' (2020) 16 *Crime, Media, Culture* 33.

<sup>32</sup> Jennifer Cobbe and Jatinder Singh, 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges' (*Preprint available on SSRN*, 4 December 2021) <<https://ssrn.com/abstract=3824736>> accessed 7 June

In sum, the prohibitions concerning manipulative AI systems may have little practical impact.

## Social scoring

The second grouping of prohibitions relate to concerns about so-called ‘social scoring’. The Act prohibits the sale or use of systems i) used by or on behalf of public authorities, ii) to generate ‘trustworthiness’ scores and which ii) lead to either unjustified or disproportionate treatment of individuals or groups, or detrimental treatment which, while justifiable and proportionate, occurs in an unrelated ‘context’ from the input data.

Trustworthiness is not defined in the Act, but can be understood as a combination of attributes that indicate that an entity will not betray another due to bad faith such as misaligned incentives, lack of care, disregard for promise-keeping (*commitment*) or through ineptitude at a task (*competence*).<sup>33</sup> Understood in this way, many scoring practices are in-scope.<sup>34</sup>

This ‘same-context’ exemption appears designed to keep reputation systems out of scope, and recalls theoretical work in privacy, including *contextual integrity*.<sup>35</sup> Yet the exemption will be difficult to operationalise.

It is unclear whether the citizen scoring characterising the ‘datafied welfare state’, commonly built with broad private sector datasets augmenting administrative data, will be in-scope.<sup>36</sup> If context is viewed narrowly, scoring can only relate to input data concerning interactions with a public authorities. A wider view however might consider credit card records and welfare support as the same context, as both involve financial flows. The Commission anticipates that a system which ‘identifies at-risk children in need of social care’ would be out of context if ‘based on insignificant or irrelevant social ‘misbehaviour’ of parents, e.g. missing a doctor’s appointment or divorce’.<sup>37</sup> The ‘European Artificial Intelligence Board’ may end up with the job of clarifying, but its guidance is only advisory.<sup>38</sup>

Public employment may also be impacted. Automated ‘social media background checks’ to score online lives seemingly concern both *competence* and *commitment* aspects of trustworthiness.<sup>39</sup> Where such systems risk detrimental outcomes, such checks would likely be prohibited due to a contextual disconnect. A public sector body using a ‘trustworthiness’-related ranking of freelancers provided to all by *LinkedIn* creates a further set of questions. Would *LinkedIn* also be liable for providing such a service to the public sector?

2021. The industry preferred term ‘democratisation’ is discussed in Sudhir Hasbe and Ryan Lippert, ‘Democratization of Machine Learning and Artificial Intelligence with Google Cloud’ (*Google Cloud Blog*, 16 November 2020) <<https://perma.cc/DL86-RJW3>> accessed 4 May 2021.

<sup>33</sup> Margaret Levi and Laura Stoker, ‘Political Trust and Trustworthiness’ (2000) 3 *Annual Review of Political Science* 475.

<sup>34</sup> This structure differs from the HLEG-AI’s initial recommendation in this area, to prohibit ‘mass scale scoring’ assessing ‘moral personality’ or ‘ethical integrity’. See High-Level Expert Group on Artificial Intelligence, ‘Ethics Guidelines for Trustworthy AI’ (April 2019) 34; High-Level Expert Group on Artificial Intelligence, ‘Policy and Investment Recommendations for Trustworthy AI’ (26 June 2019) 20.

<sup>35</sup> See generally Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford University Press 2010).

<sup>36</sup> For example, the Mosaic dataset of Experian, a consumer credit reporting company, encompasses a ‘broad and accurate range of demographic, socio-economic and behavioural characteristics on each adult and household. See Lina Dencik and others, ‘Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services’ (*Data Justice Lab, Cardiff University*, 2018) 92–93 <<https://perma.cc/39CY-H8L7>> accessed 21 August 2020; see generally across the EU, Algorithm Watch, ‘Automating Society Report 2020’ (October 2020) <<https://automatingsociety.algorithmwatch.org/>> accessed 20 June 2021.

<sup>37</sup> See, from DG CONNECT, Mazzini (n 17).

<sup>38</sup> AI Act, art 58(c).

<sup>39</sup> Miranda Bogen and Aaron Rieke, *Help Wanted - An Exploration of Hiring Algorithms, Equity and Bias* (Upturn 2018) 38–39.

What counts as a system ‘leading to’ an outcome is also unclear. Vendors will disavow negative outcomes and blame them on their users. Users (who are the authorities, rather than the citizens<sup>40</sup>) will claim that scoring was never a fully determinative factor. The result appears to be no entity clearly liable at all.

The logic behind restricting this prohibition to the public sector remains unclear. So-called AI firms control crucial infrastructures, such as delivery, telecommunications or transport. Exclusion can bring individuals grave socioeconomic consequences similar to the exclusion of state-provided services.<sup>41</sup> As with manipulation, the EU legislator has some work to do to make this provision clearly applicable to anything.

## **Biometric systems**

The Act bans some *uses* of ‘real-time’ biometric systems in publicly accessible spaces by law enforcement. An example of such a system would be a large-scale CCTV network coupled with facial recognition software. Law enforcement use of biometric identification is regulated in the Law Enforcement Directive,<sup>42</sup> which is the GDPR-type instrument for the police and similar.<sup>43</sup> Systems such as facial recognition have been easier to authorise for law enforcement purposes than other uses, such as that for a company's interest, which typically fall under the GDPR. The proposed strengthening would make the AI Act *lex specialis* to the Law Enforcement Directive, with this provision based upon TFEU Article 16 rather than 114 as the rest of the Act is.<sup>44</sup>

The Act enables Member States to authorise certain uses that fall within an exhaustive list of exceptions if accompanied by certain safeguards. Roughly summarised, the exemptions are:

- a ‘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’; and
- the ‘detection, localisation, identification or prosecution’ of a perpetrator or suspect of a crime with a maximum sentence of at least 3 years that would allow for the issuing of a European Arrest Warrant.

We can observe, firstly, that unlike the above prohibitions, this provision would allow such biometric systems to be ‘placed on the market’, meaning EU vendors can sell

<sup>40</sup> AI Act, art 3(4).

<sup>41</sup> K Sabeel Rahman, ‘The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept’ (2017–18) 39 *Cardozo L Rev* 1621.

<sup>42</sup> Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119/89 (‘Law Enforcement Directive’).

<sup>43</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119/1.

<sup>44</sup> AI Act, recital 23; see further Theodore Christakis and Mathias Becuywe, ‘Pre-Market Requirements, Prior Authorisation and Lex Specialis: Novelties and Logic in the Facial Recognition-Related Provisions of the Draft AI Regulation’ (*European Law Blog*, 5 April 2021) <<https://europeanlawblog.eu/2021/05/04/pre-market-requirements-prior-authorisation-and-lex-specialis-novelties-and-logic-in-the-facial-recognition-related-provisions-of-the-draft-ai-regulation/>> accessed 6 May 2021. National courts have not interpreted the LED as prohibiting facial recognition in its entirety, see e.g. *R (on the Application of Bridges) v South Wales Police* [2020] EWCA Civ 1058 (a case relating to pre-Brexit facts).

biometric systems which would be illegal to use in the EU to oppressive regimes.<sup>45</sup> Examples of such practices are the French firm Idemia/Morpho selling facial recognition to the Shanghai Public Security Bureau, or the Dutch firm Noldus selling facial expression analysis tool ‘FaceReader’ to the Chinese Ministry of Public Security.<sup>46</sup>

Secondly, only ‘real-time’ systems that capture, compare, and identify ‘instantaneously, near-instantaneously or in any event without a significant delay’ are prohibited. This excludes ‘post’ systems which, for example, biometrically analyse footage after an event, for example to identify individuals at protests after-the-fact,<sup>47</sup> and systems that categorise individuals biometrically.<sup>48</sup>

Thirdly, the prohibition does not ban actors from using remote biometric identification for non-law enforcement purposes, such as crowd control or public health. These uses typically fall under the GDPR. Roughly summarised, in the absence of a proportionate Member State law authorising such biometrics, the GDPR places a requirement of high-quality, individual consent for each scanned person which is effectively impossible to fulfil.<sup>49</sup>

This provision also introduces pre-authorisation familiar from state surveillance law. Competent authorities’ ‘individual use’ of a biometric system must be pre-authorised by a judicial authority or independent administrative authority (or in an emergency, shortly afterwards).<sup>50</sup> Analogous CJEU case-law regarding data retention indicates authorising bodies must have a ‘neutral stance’, notably excluding public prosecutors.<sup>51</sup> The AI Act also requires the decision of this body to be final, whereas in some Member States the executive can ignore similar bodies.<sup>52</sup>

What constitutes an ‘individual use’ to be authorised is unclear. In signals intelligence, controversial warrants can be *thematic*, relating to broad organisations, places or purposes.<sup>53</sup> In the AI Act, it is unclear if ‘individual’ could be an individual *purpose*, e.g. authorising biometrics relating to all those on a missing children list or subject to a European Arrest Warrant. As the Regulation does not explicitly require

<sup>45</sup> Some of these sales may be regulated or require transparency or authorisation under Regulation (EU) 2021/821 of the European Parliament and of the Council of 20 May 2021 setting up a Union regime for the control of exports, brokering, technical assistance, transit and transfer of dual-use items (recast), OJ L 206/1 (‘Dual Use Regulation’); which has been criticised by civil society groups in relation to a lack of prohibitions, see Access Now and others, ‘New EU Dual Use Regulation Agreement “a Missed Opportunity”’ (25 March 2021) <<https://perma.cc/P49G-P3ZR>> accessed 21 June 2021.

<sup>46</sup> Amnesty International, ‘Out of Control: Failing EU Laws for Digital Surveillance Export’ (September 2020) <<https://perma.cc/2GU5-84ZT>> accessed 21 June 2021.

<sup>47</sup> See further European Data Protection Board and European Data Protection Supervisor, ‘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) para 31. These are however included in the Title III regime, discussed below.

<sup>48</sup> Categorisation systems fall under Title IV and have weak transparency requirements, but even these have law enforcement exceptions. They could be generically added to Title III (Annex II) under delegated legislation, as the area includes biometric categorisation; regarding law enforcement it may fall under Annex II, paras 6(f–g).

<sup>49</sup> These are the most relevant conditions in GDPR, art 9; others may apply but only in extremely unusual situations. See generally European Data Protection Board, ‘Guidelines 3/2019 on Processing of Personal Data through Video Devices (Version 2.0)’ (EDPB, 29 January 2020) <[https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-32019-processing-personal-data-through-video\\_en](https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-32019-processing-personal-data-through-video_en)> accessed 6 May 2021.

<sup>50</sup> Such a system resembles national intelligence structures and the role of, for example, the *Toetsingscommissie Inzet Bevoegdigheden* in the Netherlands, or the Investigatory Powers Commissioner in the United Kingdom.

<sup>51</sup> Case C-746/18 *HK v Prokuratuur* ECLI:EU:C:2021:152 [54]. The French *parquet* is an example of a body that currently authorises such surveillance but may not be allowed to in these cases.

<sup>52</sup> Such as the French *Commission Nationale de Contrôle des Techniques de Renseignement*, which can express disapproval but not overrule the Prime Minister.

<sup>53</sup> See e.g. those avowed by the UK in the atmosphere of post-Snowden scrutiny at Intelligence and Security Committee of Parliament, *Privacy and Security: A Modern and Transparent Legal Framework* (2015) 111.

transparency over the number and type of authorisations issued, public scrutiny may be challenging.<sup>54</sup>

Either way, any authorisation of biometrics necessitates installing re-purposable infrastructure. Many already argue the AI Act legitimises rather than prohibits population-scale surveillance. The European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS) ‘call for a general ban on any use of AI for an automated recognition of human features in publicly accessible spaces [...] in any context’.<sup>55</sup> Many NGOs have come out similarly.<sup>56</sup>

### 3. Title III Regime: High-Risk Systems

The Title III regime governs AI system that pose ‘high-risk’ to ‘health, safety and fundamental rights’<sup>57</sup> in number of defined applications, products and sectors. The regime is based on and entwined with the *New Legislative Framework* (NLF) (the *New Approach* when introduced in 1985), a common EU approach to the regulation of certain products such as lifts, medical devices, personal protective equipment and toys.<sup>58</sup>

#### Scope

While the AI Act as a whole applies to all ‘AI systems’, Title III, on high-risk AI systems, only applies to two sub-categories of AI systems:

Firstly, AI systems that are products or safety components (broadly construed) of products already covered by certain Union health and safety harmonisation legislation (such as toys, machinery, lifts, or medical devices).<sup>59</sup>

Secondly, ‘standalone’ AI systems specified in an annex for use in eight fixed areas:<sup>60</sup>

- biometric identification and categorisation (both ‘remote’, as in Title II above, and applied ‘post’ the event);
- management and operation of critical infrastructure;
- educational and vocational training;
- employment, worker management and access to self-employment;
- access to and enjoyment of essential services and benefits;
- law enforcement;
- migration, asylum and border management;
- administration of justice and democracy.

<sup>54</sup> As required, for example, by the UK’s Investigatory Powers Act 2016, s 234(2)(d). Transparency of surveillance regimes differ across countries, see e.g. in the Dutch context Quirine Eijkman and others, ‘Dutch National Security Reform Under Review: Sufficient Checks and Balances in the Intelligence and Security Services Act 2017?’ (IViR, *University of Amsterdam*, March 2018) 40–41 <<https://perma.cc/LJ4Y-ZQRQ>> accessed 21 June 2021.

<sup>55</sup> European Data Protection Board and European Data Protection Supervisor (n 48).

<sup>56</sup> For example, over 60 NGOs are running a campaign ‘Reclaim Your Face’ at <https://reclaimyourface.eu/>.

<sup>57</sup> AI Act, recital 43, art 7(2).

<sup>58</sup> Little is actually new about the NLF. See Harm Schepel, *The Constitution of Private Governance: Product Standards in the Regulation of Integrating Markets* (Hart 2005) 64 (stating ‘[t]he ‘New Approach’ will most likely stay ‘new’ forever, but was not so ‘new’ even when it was launched’).

<sup>59</sup> See the list in AI Act, Annex II. Section A lists other ‘New Approach’ legislation; section B legislation is older-style product safety legislation (with a stronger role for public bodies and more detailed requirements in law) which are instead amended by Title XII to introduce new AI Act-related considerations for future delegated acts in those areas.

<sup>60</sup> AI Act, Annex III.

The Commission can, subject to Parliament or Council veto, add *sub-areas* within these areas if the application poses similar risk to an existing in-scope application, but cannot add new areas entirely.<sup>61</sup>

## The AI Act in the context of the New Legislative Framework (NLF)

Under NLF regimes, a manufacturer must undertake pre-marketing controls undertaken to establish products' safety and performance, through *conformity assessment* to certain *essential requirements* laid out in law. Manufacturers then mark conforming products with 'CE'; marked products enjoy EU freedom of movement. The philosophy of the NLF is that '[t]he manufacturer, having detailed knowledge of the design and production process, is best placed to carry out the complete conformity assessment procedure. Conformity assessment should therefore remain the obligation of the manufacturer alone.'<sup>62</sup> This distinguishes NLF regimes (including the AI Act) from pharmaceutical regulation, where a public authority (e.g. the European Medicines Agency) carries out an assessment *themselves* before granting pre-marketing approval.<sup>63</sup>

### Essential requirements and obligations

The Act contains an extensive list of essential requirements (Chapter 2) which connects to obligations of regulated actors (Chapter 3). The vast majority of all obligations fall on the 'provider: in short, person or 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.'<sup>64</sup>

Providers of high-risk AI systems must create a *quality management system*,<sup>65</sup> a standardised practice already widely present in firms.<sup>66</sup> The AI Act specifies what this entails, featuring a documented *risk management system* updated throughout the system's lifetime.<sup>67</sup>

Datasets to train AI systems must meet *data quality criteria*, including in relation to relevance, representativeness, accuracy, completeness, and application-area specific properties. Despite some requirements seeming steep — datasets being 'free of errors and complete'<sup>68</sup>, which they often are far from<sup>69</sup> — datasets only need to meet these potentially steep requirements 'sufficiently' and 'in view of the intended purpose of the system'.<sup>70</sup>

<sup>61</sup> AI Act, arts 7, 73.

<sup>62</sup> Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC OJ L 218/82.

<sup>63</sup> Note however that the AI Act is unusual in proposing that for relevant biometric systems intended to be put into service by law enforcement, immigration or asylum authorities or EU institutions, conformity assessment does require a public body, who in practice will be a data protection authority, supervising agency of those authorities, or the EDPS. See AI Act, arts 43(1), 63(5).

<sup>64</sup> AI Act, art 3(2).

<sup>65</sup> AI Act, art 17.

<sup>66</sup> The ISO 9000 series by far the best-known standard, with over 1m companies certified to ISO 9001 globally (36,000 in the IT sector), and iterative versions of the specialist implementation for software providers, now ISO/IEC/IEEE 90003:2018, available since 1997. See 2019 data obtained from International Organization for Standardization, 'ISO Survey' (no date) <<https://www.iso.org/the-iso-survey.html>> accessed 22 June 2021.

<sup>67</sup> AI Act, art 9. For providers that do not train models (e.g. expert systems or re-configured, pre-trained models), appropriate equivalents apply.

<sup>68</sup> AI Act, art 10(3).

<sup>69</sup> Curtis G Northcutt and others, 'Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks' [2021] arXiv:210314749 [cs, stat].

<sup>70</sup> AI Act, recital 44.

Much attention has been paid to the potential for AI systems to facilitate indirect discrimination, in principle illegal under EU law.<sup>71</sup> It is difficult to detect this potential unless providers know the relevant protected (and often sensitive) characteristics of affected individuals and communities.<sup>72</sup> However, the GDPR restricts the use of ethnicity data and similar sensitive data, with no specific EU-level exemption for bias detection.<sup>73</sup> The Act provides such an exemption.<sup>74</sup> The exemption can only be used in relation to high-risk systems, and only by those systems' providers. This leaves non-high-risk providers unable to rely on it.<sup>75</sup> The exemption does not provide a route for upstream data brokers to collect sensitive data on others' behalf, or to later sell to high-risk providers.

The Act contains obligations concerning the *accuracy, robustness and cybersecurity* of systems themselves, with particular regard to discrimination as systems learn,<sup>76</sup> and adversarial machine learning.<sup>77</sup> There is no explicit discussion of leakage of training data or other personal data from models.<sup>78</sup>

Providers must create *technical documentation* in line with a (Commission-amendable) Annex. The requirements are extensive; we refer the reader to them. The provider does not have to publish the technical documentation or provide it except to organisations involved in regulation or conformity assessment. However, separate provisions indicate what information must be provided as a form of *user transparency*, and what information must be *registered in a public database*. In Table 1, we abstract and group the (most salient) categories of information to be provided.

Providers must facilitate *logging* to allow traceability appropriate to a system's risks. For biometric systems, logging must include periods of use; the reference database used; and any input data leading to a match. Providers must implement a mechanism to record the identities of the 'two natural persons' checking a biometric match before it is used, and instruct a user to only use it with such a check.<sup>79</sup> Providers must only keep logs (for an appropriate amount of time) 'to the extent such logs are under their control',<sup>80</sup> else the user must instead.<sup>81</sup>

Providers must build for *human oversight*, incorporating 'human-machine interface tools' to ensure systems 'can be effectively overseen by natural persons'.<sup>82</sup> In data protection

<sup>71</sup> See, for an introduction to EU non-discrimination law applied to AI, Frederik Zuiderveen Borgesius, 'Price Discrimination, Algorithmic Decision-Making, and European Non-Discrimination Law' (2020) 31 European Business Law Review 401.

<sup>72</sup> Michael Veale and Reuben Binns, 'Fairer Machine Learning in the Real World: Mitigating Discrimination without Collecting Sensitive Data' (2017) 4 Big Data & Society 205395171774353; Kenneth Holstein and others, 'Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?' in (ACM 2019) Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2019); McKane Andrus and others, 'What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness' in (ACM 2021) Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 249.

<sup>73</sup> GDPR, art 9.

<sup>74</sup> AI Act, art 10(5). The exemption is based on strict necessity and subject to certain safeguards.

<sup>75</sup> This is despite the Act later encouraging codes of conduct to apply essential requirements to all systems. See AI Act, art 69.

<sup>76</sup> AI Act, art 15(3); see further Kristian Lum and William Isaac, 'To Predict and Serve?' (2016) 13 Significance 14; Danielle Ensign and others, 'Runaway Feedback Loops in Predictive Policing' in *Conference on Fairness, Accountability and Transparency (FAT\* 2017)* (PMLR 2018).

<sup>77</sup> AI Act, art 15(4); see generally Battista Biggio and Fabio Roli, 'Wild Patterns: Ten Years after the Rise of Adversarial Machine Learning' (2018) 84 Pattern Recognition 317.

<sup>78</sup> cf in relation to European law, Michael Veale and others, 'Algorithms that Remember: Model Inversion Attacks and Data Protection Law' (2018) 376 Phil Trans R Soc A 20180083.

<sup>79</sup> AI Act, arts 12(4), 14(5).

<sup>80</sup> AI Act, art 20(1).

<sup>81</sup> AI Act, art 29(5).

<sup>82</sup> See generally Kori Inkpen and others, 'Where is the Human?: Bridging the Gap Between AI and HCI' in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI EA '19, New York, NY, USA, ACM 2019).

law, human oversight typically relates to human dignity.<sup>83</sup> In the AI Act, human oversight instead relates to minimising risks to health, safety and fundamental rights.<sup>84</sup> A ‘four-eyes’ principle requires biometric identification systems to be designed so that two natural persons can sign off on any identification and have their identities logged, and for instructions to specify that they must.<sup>85</sup>

	Public	Users	Documentation
identity; contact details			(assumed)
member states in use		(available publicly)	(available publicly)
purpose			
conformity assessment information			
relevant standards			
instructions for use			
human oversight & technologies			
accuracy, robustness, cybersecurity	"level of"; metrics (accuracy)		metrics; test logs; test reports
risky use circumstances			"detailed information"
performance on persons/groups			"detailed information"
input data		"where appropriate, specifications"	datasheets incl. training datasets and main characteristics; provenance; labelling procedures; data cleaning
pre-determined changes			"detailed description"; techniques to ensure "continuous compliance"
lifecycle information		expected lifetime; maintenance info	"description of any change made to the system"
post-market monitoring			"detailed description [of plan]"
risk management system			"detailed description"
design specifications			"general logic"; key choices and assumptions; optimisation function; trade-off decisions; description of hardware and interacting systems
methods and steps of development			role of pre-trained models/tools; computational resources used; training methodologies

**Table 1:** Main categories of information **provided** (or **partially**, or **not**) to the public, to users, and kept by providers in technical documentation. Not fully exhaustive and grouped for comparison; refer to the Act for full information.

Somewhat strangely, no obligations for human oversight flow directly from the Act to a user. Users must simply follow the instruction manual, tying human oversight to the risk appetite of the more directly regulated provider.<sup>86</sup>

Interestingly, a leaked version of the AI Act required providers to specify *organisational measures*, notably similar to data protection guidelines,<sup>87</sup> including to ensure that overseers ‘can decide not to use the high-risk AI system or its outputs in any particular situation without any reason to fear negative consequences’, and obliged

<sup>83</sup> Meg Leta Jones, ‘The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood’ (2017) 47 Soc Stud Sci 216.

<sup>84</sup> AI Act, art 14(2).

<sup>85</sup> AI Act, arts 14(5), 12(4)(d); the Commission’s name for this is seemingly found only in slides, see Mazzini (n 16).

<sup>86</sup> AI Act, art 29(1).

<sup>87</sup> See Article 29 Working Party, ‘Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679 (WP251rev.01)’ (6 February 2018) (stating that those overseeing the decisions must have the ‘authority and competence’ to do so); see Michael Veale and Lilian Edwards, ‘Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling’ (2018) 34 Computer Law & Security Review 398, 401 (on why this is an organisational matter).

users to follow these.<sup>88</sup> In a *volte-face*, the final proposal instead emphasises the ‘user’s discretion in organising its own resources and activities for the purpose of implementing the human oversight measures indicated by the provider.’<sup>89</sup> Statements about the need for ‘competence, training and authority’ only make the recitals.<sup>90</sup>

## Conformity Assessment

These requirements are applied to providers as they must undergo *conformity assessment*. To understand conformity assessment on-the-ground, we need to explain two other actors in the AI Act: standardisation organisations and notified bodies.

### *Harmonised standards & European Standardisation Organisations*

Arguably the most important actors in the AI Act are the double-act of CEN (European Committee for Standardisation) and CENELEC (European Committee for Electrotechnical Standardisation). This may surprise readers; neither are mentioned in the text. These are two of three European Standardisation Organisations (ESOs) that the Commission can mandate to develop *harmonised standards*.<sup>91</sup>

Following a mandate, if these organisations adopt a standard relating to the AI Act, providers can follow this standard, rather than interpreting the *essential requirements*. If following the standard, providers enjoy a presumption of conformity.<sup>92</sup>

Standards can cover a legal instrument’s entire scope, or only specialist areas. For instance, the essential requirements of the Toy Directive for trampolines can be fulfilled through EN 71-14:2018; kids’ chemistry sets through EN 71-4:2013; and ‘olfactory board games, gustative games, and cosmetic kits’ through EN 71-13:2014. Standards are not free — copyright is owned by national standards bodies, and each usually costs a few hundred Euros to purchase. The Commission anticipates that the AI Act standards (it is not clear if general or specific) will first appear in the EU’s Official Journal in 2024–5, aligned with when the Regulation would be applicable.<sup>93</sup> After the industrial lobbying common in standards bodies, some aspects may look quite different from the essential requirements.<sup>94</sup>

In theory, providers do not have to follow such harmonised standards. Instead, providers could interpret the Act’s essential requirements for themselves.<sup>95</sup> This is

<sup>88</sup> Leaked AI Act, arts 11(3)(e), 18(2). The leak, dated in January, was first made available by policy subscription service POLITICO Pro (link unavailable), and republished in Natasha Lomas, ‘EU Plan for Risk-Based AI Rules to Set Fines as High as 4% of Global Turnover, per Leaked Draft’ (*TechCrunch*, 14 April 2021) <<https://techcrunch.com/2021/04/14/eu-plan-for-risk-based-ai-rules-to-set-fines-as-high-as-4-of-global-turnover-per-leaked-draft/>> accessed 1 July 2021.

<sup>89</sup> AI Act, art 29(2).

<sup>90</sup> AI Act, recital 48.

<sup>91</sup> Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council OJ L 316/12, Annex I. The last body is the European Telecommunications Standards Institute (ETSI). While the Commission has the ability to choose any of the three, recent presentations indicate they will mandate CEN/CENELEC. See the presentation by DG CONNECT, Anne-Marie Sassen, ‘Introductie van Het Voorstel Voor AI Regulering’ (Europese wetgeving Artificialiële Intelligentie, Webinar (Considerati), 15 June 2021).

<sup>92</sup> AI Act, art 40. A similar provision allows the Commission to instead propose ‘common specifications’ to specify Chapter 2 essential requirements; the main difference to harmonised standards is that failure to apply must be justified; yet the Commission has not alluded to a desire to use this, so we do not cover it extensively.

<sup>93</sup> European Commission, ‘Impact Assessment Accompanying the 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)’ (2021) 57.

<sup>94</sup> See, in relation to harmonised standards, Michelle P Egan, *Constructing a European Market: Standards, Regulation, and Governance* (Oxford University Press 2001) ch 8.

<sup>95</sup> However, if the Commission adopts common specifications, providers must justify why their measures are ‘equivalent’ to those further specified provisions. See AI Act, art 41(4).

easier said than done. Harmonised standards are both cheaper for producers, and a safer bet.<sup>96</sup> They are not as voluntary as the Commission argues. Essential requirements are often not realistically suitable for direct application.<sup>97</sup> Harmonised standards often function as a necessary point of reference for compliance through essential requirements.<sup>98</sup> In the AI Act, the requirement to consult harmonised standards is explicit.<sup>99</sup> Consequently, standardisation is arguably where the real rule-making in the AI Act will occur.

### *Controversies of harmonised standards*

The practice of delegating rule-making to bodies governed by private law such as CEN/CENELEC is controversial and sits on increasingly shaky legal ground.

Firstly, outside the field of AI regulation, it has long been argued that ‘there are structural reasons why the [New Legislative Framework] might serve the European consumer ill’.<sup>100</sup> Under-resourced consumer organisations struggle to participate in arcane private standardisation processes,<sup>101</sup> yet the outputs are important standards Member States *must* recognise. In the case of the AI Act many rights and freedoms are at stake. It is unclear whether limited existing efforts to include stakeholder representation will enable the deep and meaningful engagement needed from affected communities.<sup>102</sup> The vast majority will have absolutely no experience of standardisation, and may lack EU-level representation.<sup>103</sup> Moreover, the European Parliament has no binding veto over harmonised standards mandated by the Commission.<sup>104</sup>

Secondly, the AI Act’s value-laden nature might plant a constitutional bomb under the New Legislative Framework. Even ‘technical’ safety standards entail value-laden choices about, for example, thresholds of acceptable risk, taken under uncertainty.<sup>105</sup> The CJEU appears to be slowly recognising private standardisation bodies mandated as *de facto* NLF rule-makers cannot be free from judicial scrutiny.<sup>106</sup> Yet the NLF constitutionally relies on them being so. If such standardisation bodies are not free from judicial scrutiny, the NLF model of harmonised standards risks classification as unlawful delegation of the Commission’s rulemaking power to private bodies.<sup>107</sup> Its

<sup>96</sup> Rob van Gestel and Hans-W Micklitz, ‘European Integration through Standardization: How Judicial Review is Breaking down the Club House of Private Standardization Bodies’ (2013) 50 *Common Market Law Review*, 157.

<sup>97</sup> Scholars deride the Commission’s claim that they are as ‘pure fiction’, see Harm Schepel, ‘Case C-171/11 *Fra.bo SpA v Deutsche Vereinigung Des Gas- Und Wasserfaches*’ (2013) 9 *European Review of Contract Law*, 192.

<sup>98</sup> Gestel and Micklitz (n 96) 176. See also *Rechtbank ’s-Gravenhage*, 31-12-2008, *LJN*: BG8465 at [4.11].

<sup>99</sup> AI Act, art 9(3) (specifying providers choosing the essential requirements path must still obtain and ‘take into account’ aspects of relevant harmonised standards).

<sup>100</sup> Andrew McGee and Stephen Weatherill, ‘The Evolution of the Single Market - Harmonisation or Liberalisation’ (1990) 53 *The Modern Law Review* 578, 585; See generally Schepel (n 58) 67.

<sup>101</sup> See generally on how these processes work, Barend van Leeuwen, *European Standardisation of Services and Its Impact on Private Law: Paradoxes of Convergence* (Hart Publishing 2017) 57 et seq.

<sup>102</sup> Regulation (EU) 1025/2012, art 5 (‘European standardisation organisations shall encourage and facilitate an appropriate representation and effective participation of all relevant stakeholders, including SMEs, consumer organisations and environmental and social stakeholders in their standardisation activities’). For earlier efforts see generally Egan (n 94).

<sup>103</sup> Gestel and Micklitz (n 96) 179. See also Schepel (n 58) 111 (‘For economic operators and other interested parties, having a stake in the national standards body is the only way to get involved in European standardisation.’). On the challenges engaging traditional civil society with the technical discourses of AI and society, see generally Seeta Peña Gangadharan and Jędrzej Niklas, ‘Decentering Technology in Discourse on Discrimination’ (2019) 22 *Information, Communication & Society* 882.

<sup>104</sup> Regulation (EU) 1025/2012, art 11.

<sup>105</sup> See generally Heather E Douglas, ‘Values and Practices’ in *Science, Policy, and the Value-Free Ideal* (University of Pittsburgh Press 2009).

<sup>106</sup> See Case C-171/11 *Fra.bo* ECLI:EU:C:2012:453. The Court did not, however, follow the AG exactly, who was firmer in her reasoning that the reason to expand jurisdiction was to avoid the consequences of *de facto* transfer of public rule-making competence to private bodies. cf Case C-171/11 *Fra.bo* ECLI:EU:C:2012:176 (Opinion of Advocate General Trstenjak) [49].

<sup>107</sup> Under the ‘*Meroni doctrine*’; see Case C-9/56 *Meroni* ECLI:EU:C:1958:7; Case C-10/56 *Meroni* ECLI:EU:C:1958:8.

novel incorporation of broad fundamental rights topics into the NLF make the AI Act spotlight this tension of legitimacy.<sup>108</sup>

In sum, the Commission's long practice of privately outsourcing complex negotiations has been controversial for years. The AI Act may trigger more attention to this constitutional problem.<sup>109</sup>

#### *Self-assessment and the (limited) role of Notified Bodies*

For some products, NLF-regulated manufacturers can affix a CE certificate after 'conformity assessment based on internal control'. In the AI Act this means that they self-assess that their quality management system, system-specific technical documentation, and post-market monitoring plan follow either the essential requirements or a relevant harmonised standard/common specification.

However, under some conditions, NLF self-assessments require approval by an independent technical organisation of the provider's choosing known as a *notified body*. Notified bodies are typically private sector certification firms. They are accredited by Member States' *notifying authorities*.<sup>110</sup> Examples of notified bodies range from giants such as the German and Austrian TÜV groups, multinationals with thousands of employees who inspect and audit in a huge number of sectors, to more specialist bodies such as the Dutch *Liftinstituut*, which certifies elevators. In theory, notified bodies are transparently listed online and subject to organisational standards.<sup>111</sup> In practice, little is known about their activities, particularly due to frequent outsourcing.<sup>112</sup>

Despite pages of the AI Act establishing a regime for AI Act-specific notified bodies, there are almost no situations where their services are required. For most standalone high-risk systems (and eventually, all such systems), providers can mark the systems as in conformity using only self-assessment.

Only listed high-risk applications within the area 'biometric identification and categorisation of natural persons' must use AI Act-specific notified bodies — (initially only remote identification systems). Once harmonised standards or common specifications covering those systems exist, only self-assessment is needed. As the Commission hopes harmonised standards will exist before the application of the Regulation,<sup>113</sup> AI Act-specific notified bodies may indeed *never* be required, even for biometric systems.

AI products or components that fall under other in-scope harmonisation instruments, such as medical devices, may also require notified bodies created under the respective regime. This applies only if the product usually requires a notified body for conformity, as not to create a loophole where AI-powered products could self-assess whereas other products could not.

Political science has shown how regulatory intermediaries such as notified bodies play important roles beyond assurance, for example in translating rules, providing know-how to targets of regulation, and providing feedback to regulators and standard-

<sup>108</sup> Even before the AI Act, scholars predicted more such challenges; see Gestel and Micklitz (n 96) 153.

<sup>109</sup> Schepel (n 97) 192.

<sup>110</sup> Jean-Pierre Galland, 'The Difficulties of Regulating Markets and Risks in Europe through Notified Bodies' (2013) 4 *Eur J Risk Reg* 365, 368–69.

<sup>111</sup> Notified bodies are published in the Official Journal and on the EU's online NANDO (New Approach Notified and Designated Organisations) database. <https://ec.europa.eu/growth/tools-databases/nando/>. They follow organisational standards including EN ISO/IEC 17000, and varying legal obligations.

<sup>112</sup> Galland (n 110) 369.

<sup>113</sup> European Commission, 'AI Act Impact Assessment' (n 93) 57.

setters.<sup>114</sup> The AI Act obliges notified bodies to participate in co-ordination activities.<sup>115</sup> However, as AI Act specific notified bodies may never exist, at least in relation to non-biometric applications, this obligation seems a little futile, and leaves big gaps in knowledge flows regarding how the AI Act is functioning on-the-ground.

In sum, the AI Act gives a large role to two private standardisation organisations. CEN and CENELEC can adopt standards relating to the AI Act; standards that AI providers will follow in practice. Notified bodies checking a provider's self-assessment may play a small role, but there are few situations where they are required.

We will come back to the issue of enforcement and oversight of conformity assessment later in the paper, as this cuts across all levels of risk. For now, we turn to the 'limited risk' group of Title IV.

## 4. Title IV: Specific Transparency Obligations

Title IV lays out three transparency obligations — two for AI users, one for AI providers — that apply to all AI systems that meet their criteria.<sup>116</sup> The Commission has no powers to alter Title IV.

### 'Bot' disclosure

Providers of AI systems intended to interact with natural persons (hereafter 'bots' for short<sup>117</sup>) must design their systems such that individuals are informed they are interacting with a bot, unless it would be contextually obvious that individuals are interacting with a bot, or if the bot use is authorised by law to prevent criminal offences.<sup>118</sup>

Bot disclosure laws are not new, although none are quite like this one. In 2018, California passed the *Bolstering Online Transparency (BOT) Act*.<sup>119</sup> The BOT Act targets individuals, making it unlawful for any person to use a bot<sup>120</sup> to interact online with a Californian intending to mislead them about its artificial identity to incentivise a purchase or influence an electoral vote without clear, conspicuous disclosure.

The European Commission's voluntary Code of Practice on Disinformation commits signatory platforms to '[e]stablish clear marking systems and rules for bots and ensure their activities cannot be confused with human interactions'.<sup>121</sup> A strengthened version is expected in Autumn 2021, linking with the AI Act to tackle broad 'inauthentic

<sup>114</sup> Kenneth W Abbott and others, 'Theorizing Regulatory Intermediaries: The RIT Model' (2017) 670 *The ANNALS of the American Academy of Political and Social Science* 14; Kira JM Matus and Michael Veale, 'Certification Systems for Machine Learning: Lessons from Sustainability' [2021] *Regulation & Governance*. Note that such feedback can be useful but can also be geared towards increasing the profitability of notified bodies by reducing audit costs and rigour. See Jean-Pierre Galland, 'Big Third-Party Certifiers and the Construction of Transnational Regulation' (2017) 670 *The ANNALS of the American Academy of Political and Social Science* 263, 274.

<sup>115</sup> AI Act, art 33(11).

<sup>116</sup> AI Act, art 52.

<sup>117</sup> While we use this as convenient shorthand, the term is fraught with definitional challenges; see generally Robert Gorwa and Douglas Guilbeault, 'Unpacking the Social Media Bot: A Typology to Guide Research and Policy' (2020) 12 *Policy & Internet* 225.

<sup>118</sup> AI Act, art 52(1). Note that the criminal prevention exemption does not apply to systems that help reporting of crime.

<sup>119</sup> California's Business & Professions Code §17940, *et seq.* It has been in force since July 2019. A proposed federal bill, the 'Bot Disclosure and Accountability Act', died in the 2019–21 Congress.

<sup>120</sup> Defined as 'an automated online account where all or substantially all of the actions or posts of that account are not the result of a person'.

<sup>121</sup> European Commission, 'Tackling Online Disinformation: A European Approach (COM/2018/236 Final)' (26 April 2018) para 3.1.1; European Commission, 'Code of Practice on Disinformation' (26 September 2018) para 5 <<https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>>.

behaviour<sup>122</sup> and potentially becoming a code of conduct under the proposed Digital Services Act.<sup>123</sup>

In the AI Act, bot disclosure liability flows to *providers*, not users or platforms. This somewhat resolves the objections of scholars who criticise bot disclosure laws and proposals because practical enforcement may force exposure of the natural person behind allegedly automated content.<sup>124</sup> Enforcement of the AI Act does not require unmasking the users; i.e. the person or body using a bot. Instead, the AI Act identifies technology providers through surveilling the market for products. Market surveillance authorities have powers to compel online intermediaries to help them, but only intermediaries facilitating the sale of infringing products – not clearly, for example, the platforms the putative ‘bot’ may be communicating through.<sup>125</sup>

However, the provider-user-speaker distinction can collapse in practice. Consider the use of an AI text-generation system such as GPT-3, a tool which extends prompts into elaborate, arbitrary length strings of text,<sup>126</sup> to generate 280 character strings for posting on Twitter. Who is the provider? GPT-3 is an API to a ‘raw’ model.<sup>127</sup> To comply, should GPT-3 always return a string that ends in “#bot”? Such an interpretation would be far from technology neutral. Should some tendency to convincingly disclose itself as a bot be embedded, through training, in the 175bn parameter model itself? This seems technically daunting for a system that can just as easily produce fake legislation as a fake news article. The Act assumes a chatbot vendor pieces together and resells a system, but APIs themselves are becoming user-friendly and intuitively configurable. To make sense, the AI Act could drop its distinction between user and provider, and think in hybrids, in the style of ‘prosumer law’ long called for in information regulation.<sup>128</sup>

## Emotion recognition and biometric categorisation disclosure

Users of an emotion recognition or a biometric categorisation system must inform exposed persons of the operation of the system, except in the case of biometric categorisation permitted by law to be used for crime prevention.<sup>129</sup>

It is unclear what this provision adds to data protection law. When emotional recognition or biometric categorisation systems process personal data, data protection law requires that users of such systems inform individuals of, *inter alia*, the existence of and purposes of such processing.<sup>130</sup> Perhaps the Commission intended to mandate clear signage, given users’ lack of interest in privacy policies?<sup>131</sup> If so, the Act’s provision appears ineffective. It is not more strongly worded than the provisions of the GDPR, and the European Data Protection Board already state that users of camera systems must state the purposes on a sign.<sup>132</sup>

<sup>122</sup> European Commission, ‘Guidance on Strengthening the Code of Practice on Disinformation (COM(2021) 262 Final)’ (26 May 2021) 12.

<sup>123</sup> European Commission, ‘Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC (COM(2020) 825 Final)’ (15 December 2020), recital 69.

<sup>124</sup> Madeline Lamo and Ryan Calo, ‘Regulating Bot Speech’ (2019) 66 UCLA L Rev 988.

<sup>125</sup> 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, OJ L 169/1 (‘Market Surveillance Regulation’), art 7(2).

<sup>126</sup> Tom B Brown and others, ‘Language Models Are Few-Shot Learners’ [2020] arXiv:200514165 [cs].

<sup>127</sup> See generally on AI-as-a-Service, Cobbe and Singh (n 32).

<sup>128</sup> See generally Ian Brown and Christopher T Marsden, *Regulating Code: Good Governance and Better Regulation in the Information Age* (MIT Press 2013).

<sup>129</sup> AI Act, art 52(b).

<sup>130</sup> GDPR, art 13. Under the GDPR, the obligations are imposed on ‘data controllers’.

<sup>131</sup> It is hard to work out what the Commission intended

<sup>132</sup> European Data Protection Board (n 49) para 116.

Perhaps this provision is for where emotional recognition or biometric classification systems do *not* process personal data? Some developers claim this, such as the Fraunhofer Institute's *Anonymous Video Analytics for Retail and Digital Signage* (AVARD) system, which relies on an unpublished assertion to that effect from the Bavarian DPA for the Private Sector.<sup>133</sup> This interpretation brings many problems.<sup>134</sup> Other DPAs, national case law based on the GDPR and scholars claim that personal data is processed in these situations.<sup>135</sup> This reasoning would see the Commission implicitly legitimising a contentious and restrictive reading of the GDPR.

Either way, arguing the main issue with emotional or biometric categorisation is a *lack of transparency* risks legitimising a practice with little-to-no scientific basis and potentially unjust societal consequences. A recent literature review concluded that, '[i]t is not possible to confidently infer happiness from a smile, anger from a scowl, or sadness from a frown, as much of current technology tries to do when applying what are mistakenly believed to be the scientific facts'.<sup>136</sup> Those claiming to detect emotion use oversimplified, questionable taxonomies; incorrectly assume universality across cultures and contexts; and risk '[taking] us back to the phrenological past' of analysing character traits from facial structures.<sup>137</sup> The Act's provisions on emotion recognition and biometric categorisation seem insufficient to mitigate the risks.

### **Synthetic content ('deep fake') disclosure**

Users of AI systems that generate or manipulate image, audio or video content that appreciably resembles 'existing persons, objects, places or other entities or events' and would falsely appear to a person to be authentic are required to disclose the artificial nature of the resulting content. Exemptions exist for legally authorised crime prevention-related purposes, or necessity to exercise freedom of expression or freedom of the arts and sciences.<sup>138</sup> The narrow definition of 'user' also exempts 'personal non-professional' activities.<sup>139</sup>

The mischief this provision tackles is difficult to identify. Convincing likenesses of existing persons may harm important facets of the self,<sup>140</sup> and already trigger some personality protection.<sup>141</sup> Disclosure may only partially assist the subject; the remedy

<sup>133</sup> See Michael Veale, 'Governing Machine Learning that Matters' (PhD, University College London 2019) 217. The report from the Bavarian DPA is on file with the lead author (LDA-1085.4-1368/17-I, dated 8 June 2017). Additional potential examples are given in Damian Clifford, 'The Legal Limits to the Monetisation of Online Emotions' (PhD, KU Leuven 2019) paras 309, 311.

<sup>134</sup> Damian George and Kento Reutimann, 'GDPR Bypass by Design? Transient Processing of Data under the GDPR' (2019) 9 *International Data Privacy Law* 14; Clifford (n 133) paras 309–311.

<sup>135</sup> *R (on the application of Edward Bridges) v The Chief Constable of South Wales Police and Secretary of State for the Home Department* [2019] EWHC 2341 (Admin) [59]; Information Commissioner's Office, 'Information Commissioner's Opinion: The Use of Live Facial Recognition Technology in Public Places' (18 June 2021) 27 <<https://ico.org.uk/media/for-organisations/documents/2619985/ico-opinion-the-use-of-lfr-in-public-places-20210618.pdf>>; Peter Alexander Earls Davis, 'Facial Detection and Smart Billboards: Analysing the "Identified" Criterion of Personal Data in the GDPR' (2020) 6 *Eur Data Prot L Rev* 365.

<sup>136</sup> Lisa Feldman Barrett and others, 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements' (2019) 20 *Psychological Science in the Public Interest* 1, 46.

<sup>137</sup> Kate Crawford, *Atlas of AI* (Yale 2021) 177–78; Luke Stark and Jesse Hoey, 'The Ethics of Emotion in Artificial Intelligence Systems' in (ACM 2021) *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

<sup>138</sup> The latter is subject to 'appropriate safeguards for the rights and freedoms of third parties'; it is not further specified what these are or who applies them.

<sup>139</sup> AI Act, art 3(4). Navigating this distinction is challenging online with the rise of personal brands and influencer marketing, many of those involved already using AI 'filters' on e.g. Snapchat or TikTok. See generally Catalina Goanta and Sofia Ranchordás, 'The Regulation of Social Media Influencers: An Introduction' in Catalina Goanta and Sofia Ranchordás (eds), *The Regulation of Social Media Influencers* (Edward Elgar Publishing 2020).

<sup>140</sup> Jacquelyn Burkell and Chandell Gosse, 'Nothing New Here: Emphasizing the Social and Cultural Context of Deepfakes' (2019) 24 *First Monday*.

<sup>141</sup> Emma Perot and Frederick Mostert, 'Fake It till You Make It: An Examination of the US and English Approaches to Persona Protection as Applied to Deepfakes on Social Media' (2020) 15 *Journal of Intellectual Property Law & Practice* 32.

seems to focus instead on protecting the risk of misled audiences. Furthermore, non-human ‘entities’ do not need persona protection. In most cases, EU law already bans misleading commercial practices likely to be covered by the Act’s scope of fake ‘persons, objects, places, or other entities or events’.<sup>142</sup> If restrictions against professional deep fakes are required, it may be better to focus enforcement with consumer protection authorities instead of the product safety regulators the Act centres upon now.

The residual mischief may instead relate to situations where misplaced beliefs of authenticity present danger. For example, AI systems that increase the resolution of images, or generate 3D models from 2D images infer the remainder. People could mistakenly regard such outputs as reliable measurement rather than inference, and such mistakes could cause harm.<sup>143</sup> Yet the Act’s disclosure obligation falls on the user, not the provider. If *users* are not aware of synthetic aspects or authenticity-related software limitations, how can they protect the safety of individuals affected by their actions? Where users *intentionally* seek to deceive others using software – perhaps producing fake evidence to dispute parking tickets – why not tackle this using conventional laws of fraud?

Perhaps this provision seeks to secure some *right to reality* grounded in fundamental rights? This seems reasonable in relation to ‘fake news’ of salient events, people, places or objects. However, the provision’s scope seems too broad. It may also apply to a business using an AI stock image generator to create a bland, original scene of a board room or customer interaction for marketing purposes.<sup>144</sup> Stock photos are rarely of the real businesses in any case, and a generator may end up cheaper, easier and more tailored. Is it reasonable to require disclosure for such synthetic scenes?

Finally, as an obligation on users, this provision raises the practical enforcement questions comparable to questions regarding bot disclosure laws. How does an enforcement body investigate putatively undisclosed deep fakes? As discussed above, it is unclear whether market surveillance authorities have powers to unmask and investigate professional users of platforms who are communicating using an AI system rather than selling one. Moreover, such authorities are unlikely to have the forensics expertise needed for investigating such communications. In sum, the ‘deep fake’ provision of the AI Act raises many questions.

## 5. Harmonisation and Pre-Emption

The AI Act aims to ‘prevent unilateral Member States actions that risk to fragment [sic] the market and to impose even higher regulatory burdens on operators developing or using AI systems’.<sup>145</sup> Where the Act’s provisions entail this ‘maximum harmonisation’, Member States’ abilities to act in that area are disabled. Member States must disapply conflicting national rules and accept compliant products on their markets.<sup>146</sup> If a provision is found to not maximally harmonise an area, or only harmonises certain areas, Member States retain competence to adopt more stringent standards. The pre-emptive effect of the AI Act could have far-reaching consequences.

<sup>142</sup> Unfair Commercial Practices Directive, art 6.

<sup>143</sup> See e.g. Konstantinos Rematas and others, ‘ShaRF: Shape-Conditioned Radiance Fields from a Single View’ in (PMLR 139 2021) Proceedings of the 38th International Conference on Machine Learning; Wenming Yang and others, ‘Deep Learning for Single Image Super-Resolution: A Brief Review’ (2019) 21 IEEE Transactions on Multimedia 3106.

<sup>144</sup> See, as a rudimentary proof-of-concept, Jaemin Cho and others, ‘X-LXMERT: Paint, Caption and Answer Questions with Multi-Modal Transformers’ [2020] arXiv:200911278 [cs].

<sup>145</sup> European Commission, ‘AI Act Impact Assessment’ (n 93) 54.

<sup>146</sup> The EU’s legislative basis of approximation of laws to improve the internal market (TFEU, art 114), on which the AI Act is based, is a ‘shared competence’. Member States are in effect only permitted to legislate in this area to the extent that the Union has not.

Characterising the extent of maximum harmonisation requires identifying the material scope of the instrument (the ‘occupied field’) and determining the nature of residual Member State competence within it.<sup>147</sup> The AI Act lays out ‘harmonised rules for the placing on the market, the putting into service and the use of [AI systems] in the Union’.<sup>148</sup> The occupied field is thus not Title III ‘high risk’ systems, but all AI systems. The AI Act defines AI systems by intersecting a functional definition of systems that ‘for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with’,<sup>149</sup> with a descriptive definition based on a wide list of technologies listed in Annex I including ‘logic’ and ‘statistical’ approaches. The broad scope might not encompass *all* software, but captures some features of most. All the Act’s obligations on providers or users relate to significantly narrower subsets of this definition. However, the ‘occupied field’ with which to examine the pre-emptive effect relates to the broadest definition.

The Act therefore has an unusual misalignment between the target of its substantive obligations (primarily high-risk systems) and its material scope (all AI systems). Normally, NLF instruments do not adjust requirements (and certainly not regimes) to products of differing risk level, but instead adjust how onerous the conformity assessment ‘modules’ are (e.g. notified bodies versus internal control).<sup>150</sup> NLF instruments do not typically harmonise areas in which they impose no requirements.<sup>151</sup> The AI Act, however, seeks to both create harmonised standards, and preclude a broad array of software from further restrictions without imposing any of its own.

The way in which the AI Act may restrict further rules on *marketing* and on *use* differ, and so we look at them in turn.

### *Marketing*

Put simply, marketing of all AI systems, not just high-risk systems, is fully harmonised by the AI Act.<sup>152</sup> If Member States wish to introduce further restrictions on the placing of *any* AI system on the market, such as to limit carbon footprint or support accessibility,<sup>153</sup> they must rely on limited exceptions in Article 114 TFEU, subject to approval by the Commission.<sup>154</sup>

<sup>147</sup> Stephen Weatherill, ‘The Fundamental Question of Minimum or Maximum Harmonisation’ in Sacha Garben and Inge Govaere (eds), *The Internal Market 2.0* (Hart Publishing 2020).

<sup>148</sup> AI Act, art 1(1).

<sup>149</sup> AI Act, art 3(1).

<sup>150</sup> Decision 768/2008/EC, Annex II.

<sup>151</sup> The remaining areas are left to TFEU 34 and 36 to govern what remaining restrictions are permitted.

<sup>152</sup> With the exception of AI systems developed or used exclusively for military purposes, and, to the criticism of the EDPS and EDPB, to international law enforcement co-operation, which they see as a loophole. See European Data Protection Board and European Data Protection Supervisor (n 47) para 14.

<sup>153</sup> See generally Roel Dobbe and Meredith Whittaker, ‘AI and Climate Change: How They’re Connected, and What We Can Do about It’ (*AI Now Institute*, 17 October 2019) <<https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c>> accessed 2 July 2021.

<sup>154</sup> For example, Member States may still be able to, with the permission of the Commission, introduce a measure relating to ‘protection of the environment or the working environment on grounds of a problem specific to that Member State’ on the basis of scientific evidence, or ‘public health’ despite maximum harmonisation, see TFEU, arts 114(5), 114(8).

## Use

The extent of the AI Act's pre-emption of national rules on use of AI systems is unclear. NLF instruments before the AI Act have never placed obligations on users after installation or first use, so few clear analogies exist.<sup>155</sup>

As a starting point, the material scope of the AI Act is not only certain aspects of use, but concerns 'harmonised rules for [...] the use of artificial intelligence systems',<sup>156</sup> and aims to '[prevent] Member States from imposing restrictions on the [...] use of AI systems, unless explicitly authorised by this Regulation.'<sup>157</sup> This appears to rule out the possibility that the AI Act is a general 'minimum harmonisation' instrument, setting a horizontal regulatory floor. Such general 'minimum harmonisation' instruments are in any case not permitted by the CJEU under TFEU Article 114.<sup>158</sup>

An alternate possibility AI Act is instead a *partial harmonisation* instrument, and only transparency rules are harmonised, leaving Member States free to legislate on other issues. The Act's scope additionally states it lays down 'harmonised transparency rules for AI systems'. Those rules are listed in Title IV (discussed above) and concern both use and provision.<sup>159</sup>

In the *Phillip Morris* case, the CJEU took an escape route along those lines. Tobacco firms tried to characterise a Directive as an illegal minimum harmonisation instrument, to limit Member State's residual authority to introduce restrictions such as plain packaging requirements. The Court foiled tobacco firms' attempts by instead interpreting the Directive as only harmonising some areas. However, the Court relied on the Directive's explicit statement in its material scope that it only harmonised 'certain' aspects of packaging and labelling.<sup>160</sup> Moreover, the Directive regularly referenced to 'aspects not regulated', and included an explicit clause allowing Member States to go further in some areas.<sup>161</sup>

The AI Act lacks all the tools the CJEU relied on to escape total maximum harmonisation. The Act does note that obligations on users of *high-risk systems* are 'without prejudice to other user obligations under Union or national law'.<sup>162</sup> But no similar provision exists applying to the Act's entire scope, which itself is broad. There is therefore legal uncertainty whether existing national algorithmic transparency

<sup>155</sup> The Commission notes that '[t]he end-user is not one of the economic operators who bear responsibilities under Union harmonisation legislation' in relation to 'any operation or transaction', although this might 'fall under another regulatory regime, in particular at national level.' See Commission Notice of the 27th July 2016 on the 'Blue Guide' on the implementation of EU products rules 2016, OJ C272/1 15. The closest NLF instruments get to placing obligations on users is if incorporating assessment of installation or assembly (e.g. a lift) or distribution (e.g. a measuring instrument) is key to assessing a product, for example a lift or delicate measuring instrument – called 'putting into service'. See Blue Guide, 22.

<sup>156</sup> AI Act, art 1(a).

<sup>157</sup> AI Act, recital 1.

<sup>158</sup> See Case C-547/14 *Phillip Morris* ECLI:EU:C:2016:325 [70]–[72] (the Court noting that if it accepted that if a Directive based on TFEU 114 allowed Member States to further legislate in a field *which the Directive had harmonised*, it would have been adopted illegally on that basis). Note however that Weatherill characterises the jurisprudence on minimum harmonisation, particularly regarding other Treaty bases, as 'a mess' and indicates that there is a 'sense that the Court is not fully aware of what it is doing'. See Weatherill (n 147).

<sup>159</sup> AI Act, art 1(c).

<sup>160</sup> Directive 2014/40/EU of the European Parliament and of the Council of 3 April 2014 on the approximation of the laws, regulations and administrative provisions of the Member States concerning the manufacture, presentation and sale of tobacco and related products and repealing Directive 2001/37/EC, OJ L 127/1 ('Tobacco Products Directive'), art 1(b) ('The objective [...] is to approximate the laws [...] concerning certain aspects of the labelling and packaging of tobacco products [...]').

<sup>161</sup> *Phillip Morris* (n 158) [74]–[84]. Tobacco Products Directive, arts 1(b), 24(b–c).

<sup>162</sup> AI Act, art 29(2).

requirements applying beyond ‘high-risk’ systems, such as public sector provisions in France, may have to be disapplied.<sup>163</sup>

Even if a partial, rather than maximum, harmonisation instrument, EU primary law creates opportunities for companies to challenge use restrictions in national law.<sup>164</sup> This is no novelty of the AI Act, and a risk in any non- or partially harmonised area.<sup>165</sup> The AI Act is primarily based on EU free movement competences rather than on fundamental rights. The CJEU finds that use limitations by Member States can constitute a measure equivalent to a quantitative restriction on trade if they directly and substantially impede access to the market.<sup>166</sup> Such measures can be justified on the basis of objective justifications or public interest requirements.<sup>167</sup> However, while many justifications are possible, the Court increasingly requires Member States to justify them terms of proportionality, fundamental rights and legal certainty.<sup>168</sup> Indeed, in areas where AI Act is directly relevant, such as AI systems used by employers to manage employees, the CJEU has used freedom of movement law (in what both labour and internal market scholars characterise as troubling misapplications<sup>169</sup>) to strike down collective bargaining efforts, such as in the controversial cases of *Viking* and *Laval*.<sup>170</sup>

Some readers may feel the EU should act to prevent fragmented rules disrupting trade of AI systems. This seems defensible for a category of ‘high risk’ systems for which requirements may be complex. However, the consequence of the AI Act may be to create a stark, arbitrary divide between high-risk systems, which are regulated, and non-high-risk systems, which Member States are effectively forbidden from regulating (or become at-risk of constant challenge). As the AI Act does little to reduce fundamental rights risks of the many systems not covered by Annex III, this ‘cliff edge’ from some rules to practically none seems difficult to justify.

## 6. Post-marketing controls and enforcement

In the last two decades, New Legislative Framework regimes have evolved to include *post*-marketing controls inspired by pharmacovigilance.<sup>171</sup> The AI Act has several components of such regimes.

The AI Act gives an important role to *market surveillance authorities* (MSAs). MSAs are public bodies with wide ranging powers to obtain information, apply administrative

<sup>163</sup> Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique; décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique, art 1. See further Lilian Edwards and Michael Veale, ‘Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?’ (2018) 16 IEEE Security & Privacy 46, 48.

<sup>164</sup> cf Case C-148/78 *Criminal proceedings against Tullio Ratti* ECLI:EU:C:1979:110.

<sup>165</sup> Case C-573/12 *Ålands Vindkraft* ECLI:EU:C:2014:2037 [57].

<sup>166</sup> The logic is that such restrictions ‘have a considerable influence on the behaviour of consumers’ and leave them with ‘limited interest in buying that product’. See e.g. Case C-110/05 *Commission v Italy* (*Italian Trailers*) ECLI:EU:C:2009:66 [56]–[57]; Case C-142/05 *Åklagaren v Percy Mickelsson and Joakim Roos* ECLI:EU:C:2009:336 [26]–[27]. Rules may fall into scope based on their potential, rather than documented reality, to impede trade. See Case C-184/96 *Commission v France* (*Foie Gras*) ECLI:EU:C:1998:495 [17].

<sup>167</sup> Case C-120/78 *Rewe-Zentral AG v Bundesmonopolverwaltung für Branntwein* (*Cassis de Dijon*) ECLI:EU:C:1979:42.

<sup>168</sup> Catherine Barnard and Okeoghene Odudu (eds), ‘Derogations, Justifications and the Four Freedoms: Is State Interest Really Protected?’ in *The Outer Limits of European Union Law* (Hart Publishing 2009) 295.

<sup>169</sup> See generally Stephen Weatherill, ‘Viking and Laval: The EU Internal Market Perspective’ in Mark Freedland and Jeremias Adams-Prassl (eds), *Viking, Laval and Beyond* (Hart Publishing 2014) alongside the other chapters in that volume.

<sup>170</sup> Case C-438/05 *Viking* ECLI:EU:C:2007:772; Case C-341/05 *Laval* ECLI:EU:C:2007:809.

<sup>171</sup> ‘Pharmacovigilance is the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other medicine-related problem’. See European Medicines Agency, ‘Pharmacovigilance’ (2015) <[https://www.ema.europa.eu/documents/leaflet/pharmacovigilance\\_en.pdf](https://www.ema.europa.eu/documents/leaflet/pharmacovigilance_en.pdf)> accessed 5 July 2021. On the impact on NLF regimes, see Christopher Hodges, ‘The Role of Authorities in Post-Marketing Safety’ in *European Regulation of Consumer Product Safety* (Oxford University Press 2005); Christopher Hodges, *Law and Corporate Behaviour: Integrating Theories of Regulation, Enforcement, Compliance and Ethics* (Hart Publishing 2015) 552.

penalties, withdraw products and oblige intermediaries, including information society services (e.g. online API providers or AI-as-a-Service marketplaces), to cease offering products, or co-operate with MSAs to mitigate their risks.<sup>172</sup>

MSAs are typically government departments or regulatory agencies.<sup>173</sup> The Commission does not foresee the ‘automatic’ creation of any bespoke national authorities,<sup>174</sup> and Member States retain discretion on which authorities will be competent for the new ‘standalone’ high risk systems in the Act. AI systems in scope because they are or are parts of products regulated by harmonised legislation in Annex II are regulated by the MSA for those instruments.<sup>175</sup> In relation to Law enforcement users and Union bodies, data protection authorities will gain MSA roles.

Penalties are the maximum of 6% of global turnover 30m EUR for breaches of the Title II prohibitions or Title III data quality requirements; for other rules, the maxima are lower.<sup>176</sup> If the infringer is a public body however, penalties are chosen by Member States.<sup>177</sup>

### *Notification Obligations and Complaints*

MSAs’ main information source is through a chain of notification obligations. Users of an AI systems must monitor it and inform providers of new risks or malfunctions.<sup>178</sup> Providers must tell the MSA if their post-marketing monitoring reveals risks or non-compliance.<sup>179</sup>

However, individuals affected by AI systems have no right to complain to an MSA in the same way that, for example, they have a right to lodge a complaint to and seek a judicial remedy against a supervisory authority under data protection law.<sup>180</sup> The AI Act creates no legal right to sue a provider or user for failures under the Act, although routes may exist for litigants to argue that standards, such as those used in the AI Act, should be considered in national tort cases.<sup>181</sup> The absence of affected individuals and communities in the Act is already criticised the European Data Protection Board and the European Data Protection Supervisor.<sup>182</sup> Collectives such as consumer groups also lack any rights, such as representative complaints possible under the GDPR.<sup>183</sup> In general guidance on MSAs, the European Commission states that Member States ‘[must] ensure that consumers and other interested parties are given an opportunity to submit complaints [and have them] followed up appropriately’.<sup>184</sup> However, EU law merely requires MSAs to handle complaints competently and to consider complaints like any other information source.<sup>185</sup>

Outside the field of NLF rules, complaint mechanisms have been pivotal in developing Union case-law where regulators are reticent to challenge the practices of

<sup>172</sup> Market Surveillance Regulation, arts 14. 7(2).

<sup>173</sup> For more details, see the regularly updated resources on European Commission, ‘The Implementation of Market Surveillance in Europe’ (*European Commission, DG GROW*, no date) <[https://ec.europa.eu/growth/single-market/goods/building-blocks/market-surveillance/organisation\\_en](https://ec.europa.eu/growth/single-market/goods/building-blocks/market-surveillance/organisation_en)> accessed 1 July 2021.

<sup>174</sup> AI Act, p. 14.

<sup>175</sup> AI Act, art 63.

<sup>176</sup> AI Act, art 71.

<sup>177</sup> AI Act, art 71(7).

<sup>178</sup> AI Act, art 29(4).

<sup>179</sup> AI Act, arts 61(1), 62(1).

<sup>180</sup> GDPR, 679, arts 77-78; Law Enforcement Directive, arts 52–53.

<sup>181</sup> van Leeuwen (n 101) 20–21.

<sup>182</sup> European Data Protection Board and European Data Protection Supervisor (n 47) para 18 (noting ‘the absence of any reference in the text to the individual affected by the AI system appears as a blind spot in the Proposal’ and the absence of any ‘rights and remedies’).

<sup>183</sup> GDPR, art 80.

<sup>184</sup> Blue Guide, 104.

<sup>185</sup> Market Surveillance Regulation, arts 11(3)(e), 11(7)(a) (stating MSAs have only loose obligations to establish ‘procedures for following up on complaints or reports on issues relating to risks or non-compliance’ and to take into account complaints when taking a ‘risk-based approach’ to decide which checks to undertake).

powerful technology firms.<sup>186</sup> As only those with obligations under the Act can challenge regulators' decisions, rather than those whose fundamental rights deployed AI systems affect, the Act lacks a bottom-up force to hold regulators to account for weak enforcement. Data protection law where affected groups *can* raise complaints is already characterised by inaction and paralysis. Enforcement of the Act therefore seems likely to play out in an even more lacklustre way than it has with the GDPR to date.<sup>187</sup>

Some (non-NLF) EU product regulation contains complaint handling obligations that the AI Act could learn from. For instance, the EU Timber Regulation ensures that the monitoring authority must accept 'substantiated concerns' from civil society and other groups, and should 'endeavour' to carry out checks on operators when in possession of these.<sup>188</sup>

To make this worse, the Act's enforcement system is set up as NLF enforcement, even though only Title III is an NLF-style regime. The Act's Title II prohibitions and Title IV transparency requirements regulate users, who are brought into scope of MSA powers through a bold interpretative expansion of the MSA Regulation to simply add 'user' to 'economic operator'.<sup>189</sup> Such expansion underestimates how different regulating users will be from normal NLF oversight.

Under the Act, MSAs are expected, among other obligations, to look for synthetic content on social networks, assess manipulative digital practices of any professional user, and scrutinise the functioning of the digital welfare state. This is *far* from product regulation. MSAs are not guaranteed to be independent of the Government, as a data protection supervisory authority must be.<sup>190</sup> Apart from that, the European Commission estimates the entire enforcement of the AI Act will only take between 1 and 25 extra full-time staff at Member State level.<sup>191</sup> These authors think this is dangerously optimistic.

#### *Database of Standalone High-Risk AI Systems*

The AI Act proposes a new, central database, managed by the Commission, for the registration of 'standalone' high-risk AI systems. This approach appears to be modelled after the database and device registration requirements in the new Medical Devices Regulations.<sup>192</sup>

The database required by the AI Act makes sense to help MSAs, who otherwise might find locating illicit AI systems difficult. The Commission further proposes to make this database public, to also help 'other people', presumably civil society or journalists,

<sup>186</sup> See e.g. Case C-362/14 *Maximilian Schrems v Data Protection Commissioner* ECLI:EU:C:2015:650.

<sup>187</sup> See European Parliament resolution of 25 March 2021 on the Commission evaluation report on the implementation of the General Data Protection Regulation two years after its application (2020/2717(RSP)) para 17 (on regulatory paralysis in data protection enforcement).

<sup>188</sup> Regulation (EU) No 995/2010 of the European Parliament and of the Council of 20 October 2010 laying down the obligations of operators who place timber and timber products on the market OJ L 295/23, recital 22,

<sup>189</sup> AI Act, art 63(1)(a).

<sup>190</sup> Charter of Fundamental Rights of the European Union, art 8(3).

<sup>191</sup> European Commission, 'AI Act Impact Assessment' (n 93), Annex 3, 25. Note, the larger data protection authorities have hundreds of staff. European Data Protection Board, 'First Overview on the Implementation of the GDPR and the Roles and Means of the National Supervisory Authorities' (Report presented to the European Parliament's Civil Liberties, Justice and Home Affairs Committee (LIBE), 26 February 2019).

<sup>192</sup> Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC OJ L 117/1 (Medical Devices Regulation), arts 28–29; Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU OJ L 117/176, arts 25–26. This database is called EUDAMED, and designed *inter alia* to track Unique Device Identifiers (UDIs) and bring the EU in line with the US FDA, following requirements in the Food and Drug Administration Amendment Act of 2007. Early data protection laws also had some notification requirements

to uncover illicit AI.<sup>193</sup> With the exception of AI systems for law enforcement, migration, asylum and border management, providers must upload electronic instructions for use of AI systems in this database.<sup>194</sup> These instructions state what users must follow to avoid liability under the Act. Yet without clear complaint rights, bottom-up enforcement on the basis of this database will be significantly hampered.

The database is an interesting innovation for users who are also providers. If a company internally develops a high-risk AI system (e.g. for hiring) and puts it into service ‘for [its] own use’,<sup>195</sup> the company is both provider and user. It must declare the system on the database, and upload instructions. This seems like an important tool for accountability, also beyond AI Act requirements, but also something firms may contest in court, claiming violations of trade secrets and the like.

Similarly, users who disregard the instructions to use an AI system ‘off-label’, or substantially modify it, also become providers and therefore must declare they have done such publicly.<sup>196</sup> However, changes to AI systems which continue to learn within parameters ‘pre-determined by the provider’ do not constitute substantial modification.<sup>197</sup> Many AI systems that are based on machine learning will fall within that exception — but should they? As with the GPT-3 example above, users of general-purpose AI-as-a-Service APIs, designed to be repurposed, changed and configured, may find themselves with conformity assessment obligations without the capacity or expertise to carry them out.

## 7. Concluding Remarks

The AI Act is a world-first attempt at horizontal regulation of AI systems. It has many sensible elements, such as differentiating requirements by risk level, introducing prohibitions, and facilitating societal scrutiny via a public database of systems.

However, the Act also has severe weaknesses. It is stitched together from 1980s product safety regulation, fundamental rights protection, surveillance and consumer protection law. We have sought to illustrate how this patchwork does not make the Act comprehensive and watertight. Indeed, these pieces and their interaction may leave the instrument making little sense and impact. The prohibitions range through the fantastical, the legitimising, and the ambiguous. The high-risk regime looks impressive at first glance. But scratching the surface finds arcane electrical standardisation bodies with no fundamental rights experience expected to write the real rules, which providers will quietly self-assess against. The transparency provisions either add little to existing law or raise more questions than answers when their implications are considered. The enforcement mechanism is a creature of product safety. The regime is expected to regulate AI *users* too, yet affected communities are provided with no mechanism for complaint or judicial redress.

The pre-emptive effect of the AI Act’s maximum harmonisation raises further questions. The Act’s poor drafting risks an extraordinarily broad scope, with the supremacy of European law restricting legitimate national attempts to manage the social impacts of AI systems’ uses in the name of free trade. The Act may disapply existing national digital fundamental rights protection. It may prevent future efforts to regulate AI’s carbon emissions or apply use restrictions to systems the Act does not

<sup>193</sup> European Commission, ‘AI Act Impact Assessment’ (n 93) 56. Databases mandated by law in fields such as trademarks have been discussed as potentially helpful to algorithmic accountability. See e.g. Amanda Levendowski, ‘Trademarks as Surveillance Transparency’ (2021) 36 Berkeley Tech L J \_\_.

<sup>194</sup> AI Act, Annex VIII, para 11.

<sup>195</sup> AI Act, art 3(11).

<sup>196</sup> AI Act, art 28.

<sup>197</sup> AI Act, art 43(4).

consider 'high-risk'. Counterintuitively, the Act may contribute to deregulation more than it raises the regulatory bar.

This paper cannot and has not covered all the Act's facets. Many aspects are omitted and deserve further scrutiny. We urge legislators and civil society to rise to this challenge, and hope that we have demonstrated some of the complexities making this a particularly important instrument to analyse throughout its legislative process.