On the path to the future

Varosanec, Ida

Published in:
International Review of Law, Computers & Technology

DOI:
10.1080/13600869.2022.2060471

Publication date:
2022

International Review of Law, Computers & Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cirl20

On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI

Ida Varošanec

To cite this article: Ida Varošanec (2022) On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI, International Review of Law, Computers & Technology, 36:2, 95-117, DOI: 10.1080/13600869.2022.2060471

To link to this article: https://doi.org/10.1080/13600869.2022.2060471

INTERNATIONAL REVIEW OF LAW, COMPUTERS & TECHNOLOGY
2022, VOL. 36, NO. 2, 95–117
https://doi.org/10.1080/13600869.2022.2060471

Routledge
Taylor & Francis Group

BILETA SPECIAL EDITION

On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI

Ida Varošanec

Faculty of Law, University of Groningen, Groningen, Netherlands

ABSTRACT

Transparency is the currency of trust. It offers clarity and certainty. This is essential when dealing with intelligent systems which are increasingly making impactful decisions. Such decisions need to be sufficiently explained. With the goal of establishing 'trustworthy AI', the European Commission has recently published a legislative proposal for AI. However, there are important gaps in this framework which have not yet been addressed. This article identifies these gaps through a systematic overview of transparency considerations therein. Since transparency is an important means to improve procedural rights, this article argues that the AI Act should contain clear transparency obligations to avoid asymmetries and enable the explainability of automated decisions to those affected by them. The transparency framework in the proposed AI Act leaves open a risk of abuse by companies because their interests do not encompass considerations of AI systems' ultimate impact on individuals. However, the dangers of keeping transparency as a value without a legal force justify further reflection when regulating AI systems in a way that aims to safeguard opposing interests. To this end, this article proposes inclusive co-regulation instead of self-regulation so that impacted individuals as well as innovators will be empowered to use and trust AI systems.

KEYWORDS

AI Act; transparency; regulation

1. Introduction

Artificial intelligence (AI) has been increasingly adopted in government decision-making to improve the quality, efficiency and effectiveness of public services and procedures. However, legal literature and courts have expressed concerns regarding the risks of automated decision-making for fundamental rights. The basis of these decisions is unclear, and can thus give rise to many issues with the rights of individuals under scrutiny. These risks are often explained by information asymmetry created by the insufficient transparency of black-box AI systems.

CONTACT Ida Varošanec i.varosanec@rug.nl https://www.linkedin.com/in/ida-varo%C5%A1anec-b2b091133/ https://twitter.com/IdaVarosanec Faculty of Law, University of Groningen, Oude Kijk in 't Jatstraat 26, 9712 EK Groningen, The Netherlands

Based on fundamental values stemming from EU Treaties and Charter, processes ought to be transparent, and decisions explainable to those affected by them. Opacity is the common denominator of automated decision-making systems. The lack of transparency and explanations mean that it is more difficult for individuals to challenge the basis of automated decisions. Parties need transparency because of the impact of these decisions on them, to identify possible errors within, and to have the possibility to contest, correct and receive compensation for erroneous decisions (Doshi-Velez and Kortz 2017). Therefore, transparency is a crucial tool in asserting responsibility, accountability and enabling the preservation of citizen rights. Although there are many calls for more transparency in AI, this is not as easy as it seems. On the one hand, transparency is a key value and a principle required by public law which *inter alia* requires public bodies to give reasons for their decisions involving civil consequences to individuals (Wade and Forsyth 2000). On the other hand, AI systems equally need opacity for a wide array of reasons of private interest such as the preservation of trade secrecy which protects commercial interests of companies developing AI systems and products (Alì and Yu 2021). In this article, transparency is seen as encompassing explainability necessary to enable the exercise of rights which it is purported to uphold.

The proliferation of AI systems in various sectors within the EU has led to a proposed new regulation – the AI Act proposal.[1] Due to the opaque nature and potential impact of AI systems, the proposed AI regulation should contain clear transparency obligations so that impacted individuals as well as innovators are empowered to use and trust AI systems. Instead, the legislator seems to have chosen another route – entrusting transparency obligations largely to self-regulated ethical frameworks; this article will explain why that approach is misguided. In observing how transparency is addressed within the evolving EU legal framework for AI, this article inquires: *How to regulate AI to promote transparency?* It argues that transparency should not be left to self-regulation. Further, it contends that the transparency framework in the proposed AI Act leaves open a risk of abuse by companies who will actively ignore transparency obligations and the need to assess the impact of their AI systems on individuals if transparency is to be addressed through self-regulation. Thus, existing impediments (e.g. proprietary rights) to transparency might be exacerbated and power/information asymmetries created.

Many knowledge gaps still exist in the area of AI governance. Since the regulatory framework for AI in the EU is at the beginning of its evolution, this article builds upon the nascent literature focusing on transparency as a crucial tool for enabling trust. It complements the existing scholarship in synthesising extant considerations applied to the current state of AI regulation in the EU. However, it differs in that it clearly identifies issues pertaining to self-regulation of AI systems by companies which can lead to power/information asymmetries, and proposes an inclusive form of regulation.

This article begins by delineating the concept of transparency, explaining why that is important *inter alia* for fostering trust and accountability in Section 2. This is followed, in Section 3, with a discussion of the approach to transparency in the proposed AI Act, and in particular, highlighting that the proposal's intent to achieve transparency through self-regulated ethical frameworks is flawed. Section 4 presents the importance of choices in modes of regulation which should be carefully considered in the context of AI and transparency, in order to avoid the aforementioned risks, while Section 5 concludes that the AI framework should address existing flaws.

2. Transparency in the context of AI

There is no universal definition of transparency. Instead, its meaning is tailored to the relevant field of application (Ball 2009). Commonly, it can signify the characteristic of being easy to see through, or the quality of openness without secrecy. As a legal term, it can be said to mean an insight into legislative, administrative or judicial proceedings, and is a part of the freedom to expression and the right to information including a right to access official documents (Reichel 2021). It is part of EU law as reflected by Article 1 of the Treaty on European Union (TEU) which requires decisions to be taken as openly as possible to the citizen.[2] For the purpose of outlining its importance when dealing with AI, it is examined and placed in the context where it should be utilised by public bodies in their interaction with individuals. It will become clearer that transparency is a fragmented notion of great importance within the context of AI systems.

2.1. A tool for rights

Pursuant to Article 296 TFEU and Article 41(2c) of the Charter of Fundamental Rights of the EU, public authorities are required to give reasons for their legal acts and decisions. The statement of reasons must be 'sufficiently clear' and unequivocal to allow the Court to review its legality and inform the persons concerned whether the decision is erroneous and enable them to challenge its validity.[3] This duty to give reasons is – aside from being a transparency obligation – thus an important means to facilitate accountability and individual access to justice (Fink 2021). However, further clarification is needed on what reasoning is required of public bodies to satisfy the 'sufficiently clear' requirement, and what is thereby required from the design of an AI system used by such authorities. Even though public law is not harmonised in the EU, the duty to give reasons is found in most of national laws (Opdebeek and De Somer 2016) and will thus depend on a proper transparency framework for AI.

Transparency is not an end, but a necessary condition for trust and exercise of procedural rights (von Eschenbach 2021). It is a means of achieving 'trustworthy' AI enabled by explainability of the decisions made by these systems. It has been shown that transparency reduces uncertainty in AI decision-making systems and increases trust (Liu 2021). Transparency is thus an important instrument to achieve a potential multitude of other fundamental values which depend on its existence (Heald 2012). This is crucial, because design flaws of algorithms or their training datasets can cause failures which can lead to undesired consequences and violations of fundamental rights (Eubanks 2018).

2.2. Explainability

Explanations of decision-making technologies in each case, as well as the choices made, would assist in asserting responsibility and liability inquiries. Humans require a narrative form of explanation which opposes the binary nature of AI systems' outputs. As noted by Reed, Grieman, and Early (2021), most citizens would not trust any AI system if they were simply told 'We cannot explain how it works, but it is really safe'. This prompted a development of an entire field of eXplainable AI (XAI) which focuses on designing tools that can enable explanations for the decisions produced by complex autonomous systems.

However, these tools cannot currently provide a unified approach to explain the decisions of all AI systems, as different systems and different stakeholders require different explainability requirements. For instance, what regulators need to know is different from what consumers need to know. This is because each group has different rights and obligations. Regulators require not only a short non-technical description of the algorithmic tool and (the reasons for) its use, but they also need details on how the tool works and the data it uses (Kingsman et al. 2021). Similarly, as imposed by many European administrative law systems, public bodies using AI systems need to observe principles of good administration and the duty to give reasons in their decision-making, thereby requiring information on what elements were considered in reaching a certain decision in order to explain those elements to the affected citizens in the justification of their decision (Opdebeek and De Somer 2016; Wade and Forsyth 2000, 472). Consumers, however, will want to know whether the product embedded with an AI system is safe to use. Transparency for them might require access to testing data and results. Consequently, a transparency framework should consider the nature and goals of the stakeholders (Weller 2019; Felzmann et al. 2019).

2.3. Accountability and limitations

Many calls for transparency advocate the 'opening of the black box' of AI systems (Wischmeyer 2020; Duval 2021; Pasquale 2016). In this sense, transparency is meant to ensure accountability and governance. It can be conveyed as a 'system of observing and knowing that promises a form of control' (Ananny and Crawford 2018). However, transparency of this kind has limits. For instance, gaining an insight into a system's inner workings is inadequate because it creates an illusion of promising meaningful accountability which transparency cannot provide (Heald 2012; Ananny and Crawford 2018, 978). On the contrary, focus on transparency can have undesirable consequences. One such example can be seen in the known practice of 'gaming', whereby those who know enough about the technology obtain goods or services unfairly (Diakopoulos 2016; Ghani 2016). Additionally, transparency obligations are at risk of causing what Stohl, Stohl, and Leonardi (2016) call 'inadvertent' or 'strategic' opacity. In both cases, important information is concealed amidst great quantities of information provided by actors complying with transparency regulations. Sifting through such an impossible amount of information renders transparency unusable.

Transparency is not enough for understanding AI systems, as it risks prioritising 'seeing' over 'understanding'. As Ananny and Crawford (2018) put it, 'without nuanced understandings of the kind of accountability that visibility is designed to create, calls for transparency can be read as false choices between complete secrecy and total openness'. Furthermore, transparency has temporal limitations, as different moments in time may produce different kinds of system accountability (Ananny and Crawford 2018, 982). Since these systems are prone to rapid changes in their self-development, transparency should contain a temporal dimension to allow observance and understanding of previous iterations of the AI system. Otherwise, even if the source code, full training and testing data were provided, it does not guarantee a complete oversight over its functionality.

AI systems' autonomous nature, and the ability to learn and self-modify without the presence of programmers, necessitate the call for oversight (Alì and Yu 2021, 4). Thus,

transparency is pivotal to gain an insight into AI systems to understand their logic and regulate their behaviour. It would enable 'accountability of automated decision-making processes, the possibility to prevent potentially harmful conducts and to correct any source of unequal, illegal or undesirable behaviour' (Alì and Yu 2021, 6). This can be done *ex ante* and *ex post*. For instance, Felzmann et al. (2019) separate transparency into *prospective* and *retrospective* transparency. The former signifies an accountability mechanism which provides an insight into information about the data processing and the working of the system upfront (Zerilli et al. 2019). The latter refers to explanations and rationales provided for auditing purposes – namely, revealing for a specific case how and why a certain decision was reached. Ideally, a transparency framework for AI would encompass both versions. However, transparency in itself cannot facilitate a complete correction of wrongs. Instead, it is an important mechanism for authorities who are tasked with protecting our rights. AI regulation is a pivotal step in enabling them to do so in a way that incites trust on the complete stakeholder spectrum.

3. Transparency in the evolving AI regulation

The need for regulation started taking root when AI suddenly became a determinant of EU competitiveness.

3.1. Mapping out the notion of transparency on the way to the EU legislative proposal of the AI Act

Although regulation of the use of automated systems and discussions involving AI regulation have not been novel in the EU[4] (See Hildebrandt and Gutwirth 2008), the first intention for this new framing materialised on 10 April 2018 when Member States agreed to cooperate on the EU strategy for AI by signing a Declaration of Cooperation on Artificial Intelligence.[5] They agreed to ensure *inter alia* an adequate legal and ethical framework, and to 'prevent harmful use' of AI systems.[6] This is the beginning of a segregation of ethical and legal frameworks.

3.1.1. Artificial intelligence for Europe

This narrative continued in further documents which set the aims of legislation – namely boosting AI uptake and ensuring an appropriate ethical and legal framework based on EU values and the Charter of Fundamental Rights.[7]

Achieving these aims requires transparency because it is a cornerstone of EU values and because it represents a tool for sharing information between various stakeholders and thus enables the pursuit of their interests or those which they aim to protect – innovation, preservation of rights, the building of relevant frameworks and safely embedding AI in the society. However, even in these inception phases, there is divergence into separate frameworks. The importance of research into the explainability of AI systems to allow humans to understand the basis of their actions has merely been mentioned in these documents. Being well-informed is crucial for exercising one's autonomy and represents a vital element of transparency (Buijze 2013, 226). In relation to individuals' decision-making capabilities, the ECtHR has explained that existential matters as a basis of the right to information encompass all domains in which they are confronted with

the need to make fundamental choices in their life (Rouvroy and Poullet 2009).[8] Thus, the Commission could have been more elaborate on the necessity of making AI processes explainable, particularly when it comes to employing such systems in situations of impact on citizens where the duties of users require the enablement of the right to know the reasons for a certain decision.

3.1.2. (Draft) AI ethics guidelines

In the meantime, AI HLEG – an expert group recruited by the Commission – created and published the first draft of Draft AI Ethics Guidelines in December 2018.[9] During the subsequent three-month consultation period, 500 comments were submitted by various stakeholders.[10] Many stakeholders requested clarification on the enforcement of the ethical guidelines and the consequences of non-adherence to them. The Commission responded that the application of guidelines by AI developers, deployers and users is purely voluntary. This is a point of concern if certain crucial considerations – such as transparency – are left to voluntary application. In such a case, developers are not only able to avoid transparency regarding the AI systems they produce, but they are also allowed to limit transparency to suit their own commercial interests, thereby diminishing an important tool for the enjoyment and preservation of fundamental values. In practice, this might mean that an affected individual is banned from asserting his rights to know the reasons for an AI decision because the AI developer is permitted to decide when an individual may or may not seek reasons for an AI decision. There is a reason why safeguards exist, and transparency is one of those safeguards. Hence, it deserves to be placed in a legal framework, instead of an unenforceable ethical framework.

Subsequently, the guidelines were amended taking into considerations remarks made and the AI HLEG group finally presented its 'Ethics Guidelines on Trustworthy AI'.[11] The guidelines specify that in case of a 'black box' obstacle, alternative measures – such as traceability, auditability and transparent communication on the capabilities of the AI system – should be in place. With regard to transparency, technical processes and the related human decisions are advised to be explainable in a way that is adapted to the stakeholder concerned – be it a layperson, a regulator, or a researcher. These guidelines aim to provide a framework for AI stakeholders to develop and deploy trustworthy AI systems in a way that secures their benefits and avoids their potential harmful consequences for society. Regardless, certain conflicts remained unaddressed – such as the one between the transparency of AI systems and intellectual property rights impeding them.

The Commission's reflection on these guidelines emphasised that certain provisions in EU law already reflect several requirements established by AI HLEG, and that ethics guidelines are needed which build upon the existing regulatory framework instead.[12] Although there are existing provisions which would capture the use of AI systems, transparency obligations are rare. AI regulation offers a substantial opportunity for improvement.

3.1.3. Working document 'policy and investment recommendations for trustworthy artificial intelligence'

The AI HLEG also emphasised the importance of public services and the need to consider the capacity of current laws to offer an appropriate scope of protection.[13] Given the scale of operation of such services, AI development and deployment ought to be transparent to ensure consistency with the principles of good administration, respect for fundamental

rights, democracy and the rule of law. Therefore, the group recommended to consider the capacity of current laws to offer *inter alia* the 'appropriate scope of intellectual property rights protection, and whether GDPR mandated transparency and explainability offer sufficient protection'.[14]

Articles 13–15 of the General Data Protection Regulation (GDPR) state that people whose interest has been affected by an algorithmic decision have the right to receive information and to have the decision explained to them.[15] However, this right contained in the GDPR is disputed in the literature and has remained ambiguous since there is no agreement on its existence in the first place (Wachter, Mittelstadt, and Floridi 2017; Casey, Farhangi, and Vogl 2019; Edwards and Veale 2017; Goodman and Flaxman 2017; Selbst and Powles 2017; Kaminski 2019). For instance, the GDPR prescribes that a data subject should be given a 'meaningful' explanation.[16] However, it remains uncertain what level of detail such an explanation should necessitate (Felzmann et al. 2019). Possible information provided to users should enable them to determine the main elements in a decision and understand how they alter the outcome (Doshi-Velez and Kortz 2017; Wachter, Mittelstadt, and Floridi 2017; Zerilli et al. 2019).[17] Unfortunately, such an explanation will not provide a justification for the reason why such a decision has been taken, nor does it shield the user from suffering the consequences connected to it (Felzmann et al. 2019). Consequently, although an important inspiration for the right to explanations for automated decisions, the GDPR is not sufficient on its own to enable universal transparency that can be applied across different contexts.

3.1.4. White Paper

Transparency also appeared in the White Paper published in February 2020 in which the European Commission emphasised the need to develop an 'ecosystem of trust'.[18] It listed transparency as one of the key requirements for AI in the EU that had not received sufficient attention when calling for improvements to be made to the legal framework. The Commission envisioned the opaqueness of AI systems being addressed by imposing specific transparency requirements. However, it is uncertain whether it envisioned transparency as a soft-law or hard-law requirement.

From the beginning stages, ethical and legal frameworks have been separated and transparency has been placed within the former which undermines its importance when dealing with AI systems that have impact on individuals' lives. Although mentioned in each document leading up to the proposal, transparency has not been defined in terms of specific requirements. If ethical framework implies soft law approach in the form of self-regulation, transparency should not feature only therein. Finally, insufficient attention has been paid to explainability and potential information asymmetries.

3.2. AI Act proposal: transparency obligations framework

On 21 April 2021, the European Commission published a proposal for a new AI regulation – the so-called Artificial Intelligence Act – to foster trust in AI technologies.[19] The proposal represents a significant step in the regulation of AI. In its substance, the proposal lays down rules as per levels of risk associated with AI systems – those that carry (1) unacceptable risk, (2) high risk, and (3) limited risk. The fourth category – that of systems that pose

a minimal risk (e.g. spam filters) – although within the material scope, are not subject to any concrete rules.

The first category – (1) unacceptable risk – concerns AI systems that are a clear threat to the safety, livelihoods and rights of persons.[20] Specifically, subject to a 'significant harm' requirement, Article 5 forbids AI systems or applications which manipulate human behaviour in circumvention of their free will (e.g. toys using voice assistance to encourage dangerous behaviour in children) and detrimental systems that enable 'social scoring' by governments when they are employed in social contexts differing from original ones where data was generated.[21] It also sets the conditions on the use of 'real-time' biometric identification systems in publicly accessible spaces. The 'significant harm' requirement which triggers prohibitions imposed by the AI Act on 'unacceptable risk' systems does not suffice in the protection of individuals' interests. As Veale and Borgesius (2021, 12–13) emphasise, this threshold does not encompass harmful consequences which occur cumulatively over time or those caused by other users *ex post*.

Throughout the development of this AI regulatory framework, transparency is posited more as a goal, rather than a concrete set of hard-law rules which would enable (legal) certainty. The AI Act proposal is an important step in defining the role we want AI to have in our societies. However, it will become clear that it still needs improvement.

3.2.1. High-risk AI systems

Title III is dedicated to (2) high-risk AI systems. Article 6 in conjunction with Annex III describes which systems are considered high risk. The bifurcated classification comprises systems (1) embedded in products which are subject to third-party assessment under sectoral legislation, and (2) those which are stand-alone, and thus not components of products, but are deemed to be high-risk when utilised in certain areas. It is envisioned that if the AI system featured in a product as a safety component which complies with the relevant sectorial legislation will also be compliant with the AI Act. Stand-alone high-risk AI systems include technology used in: critical infrastructures (e.g. transport), that could jeopardise the life and health of citizens; educational/vocational training, that may determine the access to education and professional course of individuals (e.g. scoring of exams); safety components of products (e.g. AI application in robot-assisted surgery); employment, workers' management and access to self-employment (e.g. CV-sorting recruitment software); essential private and public services (e.g. loan scoring); law enforcement that may interfere with fundamental right (e.g. evaluation of the reliability of evidence); migration, asylum and border control (e.g. verification of authenticity of travel documents); and administration of justice and democratic processes (e.g. applying law to a concrete set of facts).[22]

High-risk AI systems are subject to several requirements before they are placed on the market. Pursuant to the AI Act proposal, they ought to entail adequate risk assessments and mitigation systems; the high quality of the datasets feeding the system in order to minimise risks and discriminatory outcomes; logging of activity to ensure traceability of results; detailed technical documentation providing all necessary information on the system and its purpose for the sake of assessment of compliance by authorities; appropriate human oversight measures to minimise risk; and high level of robustness, security and accuracy.[23] Quality of data requires the training, validation and testing data to be relevant, representative, free of errors and complete.[24] However, no data can be error-free. Quality

of data will not guarantee a lack of discrimination and bias. Moreover, it is unclear whether these requirements will enable adequate transparency for different actors. These elements are missing from the proposal.

Transparency obligations and provision of information to users are imposed on providers.[25] Developers are thus instructed to 'Design and develop AI systems in a way that ensures that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately', provide instructions for use that include 'concise, complete, correct and clear information that is relevant, accessible and comprehensible to users'.[26] However, it is uncertain what information would afford sufficient transparency and who will decide whether it would enable users to interpret and use an AI system's output appropriately.

Providers are required to compose technical documentation which contains information necessary to assess their compliance with other requirements.[27] For instance, a detailed description of the elements of an AI including *inter alia* the design specifications of the system and system architecture ought to be included.[28] The proposal also imposes record-keeping requirements aimed at providing traceability. It remains unclear if this is enough for transparency we want and need as citizens with rights. The proposal seems to envision a solution in setting an obligation of human oversight for high-risk AI systems. Natural persons overseeing the use of AI systems should be able to fully understand the capacities and limitations of the system; remain aware of potential automation biases in (over)relying on decision-making outputs of an AI system; and decide or intervene in the operation thereof when safety and fundamental rights are at risk.[29]

Providers are expected to comply with the aforementioned requirements by means of a self-assessment procedure. Once they do, they can complete the EU declaration of conformity and affix the *Conformité Européenne* [CE] marking to enter the EU market.[30] This latter aspect is highly problematic, as it leaves space for abuse by companies who pursue different priorities and have different interests.

Users are only required to follow an instruction manual provided by the provider.[31] However, information to be provided to users does not necessarily mean that individuals as third persons are afforded necessary safeguards in the form of transparency. The requirement of national authorities to give reasons to justify decisions affecting EU rights closely relates this right with the duty to give reasons (Tridimas 2006, 445).[32] This means that legal and factual features as well as other considerations which led to the decisions need to be stated.[33] However, the AI Act proposal imposes strict confidentiality standards.[34] Unfortunately, by allowing for self-regulation practices which are then mandatorily transferred onto the users – such as public authorities – in the form of 'instructions for use', the proposal *de facto* ties and thus imposes reliance of the public sector on the private sector which holds the power over public authorities by setting rules for how those authorities should use AI systems. Certainly, the proposal did not account for issues that can arise in cases of obstacles preventing compliance with this requirement. If we wish to manage how organisations and public authorities use AI systems, ethics cannot substitute hard law (Black and Murray 2019, 15).

3.2.2. Non-high-risk AI systems

Non-high-risk AI systems comprise those that pose a (1) 'minimal risk' and those of a (2) 'limited risk'. As the name implies, minimal risk AI systems (1) encompass 'the vast

majority of AI systems currently used in the EU' which are of little or no risk (e.g. spam filtration). Article 69 of the proposed AI Act might shape their development by encouraging codes of conduct to be made and used. It appears that this provision is an attempt to encourage (rather than oblige) companies to bring the level of caution to the standards required for high-risk AI systems in terms of transparency, human oversight and robustness (Gaumond 2021). If so, the insufficient standards as identified in the previous section do not provide much confidence in compliance.

The category of *limited risk* features AI systems which pose issues in terms of transparency and thus require special disclosure obligations – specifically, deep fakes, systems intended to interact with people, and emotion recognition and biometric categorisation systems.[35] Providers are required to design and develop AI systems in such a way that natural persons are informed that they are interacting with an AI system.[36] Similarly, users of an emotion recognition or a biometric categorisation system are to inform natural persons exposed about its operation. Finally, artificial generation or manipulation in the case of 'deep fakes' should be disclosed. Overall, users of limited-risk AI systems need to be transparent about their artificial nature towards the natural persons exposed to it.[37] This is not sufficient in cases where such systems potentially have high-risk consequences.

The proposed regulation heavily regulates certain AI systems which carry high levels of risk. However, uses that are of limited or no risk are left to soft law instruments like codes of conduct. This is problematic for several reasons. It is concerning because certain systems on minimal risk can quickly turn into something more devious. Existing regulatory systems do not clearly capture the use of AI, especially when it comes to stand-alone systems. Additionally, there is a risk of patchwork regulation encompassing overlapping or lacking norms if we rely on existing regimes. For instance, since minimal risk systems lack rules but are within the material scope of the AI Act, they are subject to regulatory pre-emption. This means that even if Member States create transparency rules for such AI systems, they are likely to be disapplied (Edwards and Veale 2017; Veale and Borgesius 2021, 90). Consequently, attention should also be paid to the reversible nature of AI systems which are perceived not to pose significant risks and are thus also left to self-regulation.

3.2.3. Conformity assessment as per 'New Legislative Framework' and 'other harmonisation legislation'

If a high-risk AI system is embedded in a product as a safety component, sectoral safety legislation will apply. Annex II differentiates between the legislation based on the 'New Legislative Framework' and the 'Old Approach legislation'.[38] In order to avoid overlapping requirements, the proposal directs high-risk AI systems towards the former for conformity assessment procedures. The main *ratio* for self-certification and conformity assessments pursuant to the New Legislative Framework rests on the fact that manufacturers are the most knowledgeable to conduct such an assessment (Veale and Borgesius 2021, 38).[39] Therefore, AI systems which are part of a product falling under one of the directives or regulations will have to be tested, inspected and certified by manufacturers prior to placing on the EU market.[40] However, it is uncertain whether manufacturers' incentives are sufficient to ensure the existence of relevant safeguards.

The AI Act is posited as a risk-based regulation which, as the name implies, seeks to control relevant risks within a regulated field. In its essence, it prioritises regulatory actions pursuant to an assessment of the risks presented to the regulatory body (Black 2005, 2010a, 2010b; Baldwin, Cave, and Lodge 2011, 282). Law is thus complemented by other forms of decision-making such as risk assessments. To avoid misconstruction of technological risks *de lege*, the regulator relies on the industry to mark the risks. Risk assessments are often presented as objective and neutral. However, empirical studies have shown that risk-based practices are always strongly connected to social relations and concepts (See Ewald 1991; Wynne 2002; Stirling 1998; Van Dijk, Gellert, and Rommetveit 2016). Consequently, it enhances the scientific and democratic aspects of risk-regulation by requiring decision-makers to use science and creating opportunities for public participation (Fisher 2010). The divergence of perceptions of risks between the regulator and the public can make risk-based regulation difficult. The regulator, the industry and citizens may thus perceive what constitutes a risk differently thereby making adaptation onerous (Baldwin, Cave, and Lodge 2011, 289). However, it is an opportunity to bring public and private actors closer together to collaborate and ensure equal account of interests in regulation. This cooperation between public and private actors is crucial for the effective regulation of AI. Risk assessments play an important role in that regard as they enable transparency through exposure. Specifically, in the AI regulation, they can be an enabler of *ex ante* transparency if construed in a way that takes rights and interests into account.

Ex ante conformity assessment is to be carried out not by external parties but the companies themselves, pursuant to standards created by the European Standardisation Organisation (ESO).[41] Providers could choose to follow them instead of having to interpret 'essential requirements' for themselves (Veale and Borgesius 2021, 51). Standards are expected to appear in 2025 – presumably around the same time as the AI Act would come into force.[42] However, self-assessment has been criticised for its unreliability, cloudiness and discretionary nature and thus the strengthening of *ex ante* obligations has been strongly advocated (Smuha et al. 2021). Delegation of rulemaking to ESOs is equally problematic since these bodies are governed by private law. The Commission's long practice of outsourcing complex rule-making negotiations to private bodies has been controversial (Veale and Borgesius 2021, 57). These are private organisations assigned with public tasks; however, they need not observe the important guarantees of public law such as transparency (Volpato 2021). To the contrary, these bodies are considered to have commercial interests and can thus require royalties for the use of their copyrighted standards.[43] This is illustrative of the existing dichotomy between public and private interests. However, a series of ethical and legal decisions must be made which cannot be outsourced to private organisations (Ebers 2021). Instead, they should undergo a legislative procedure which can be shaped by the stakeholders and which afford them the value proportionate to interests at stake.

Generally, where risks for individuals exist and can be compensated for, regulation comprises *ex post* liability regimes which may or may not be complemented by oversight and enforcement (Black and Murray 2019, 14–15). However, levels of risk dictate the breadth of *ex ante* as well as *ex post* requirements. Established frameworks for products (which AI is only becoming part of) have their own mandatory compliance requirements which are proportionate to the risks posed by these products. This cannot be said to be captured by the new AI Act proposal due to asymmetries to which we now turn.

3.3. The information asymmetry

Achieving transparency on a technical level in the operationalisation of an AI system is difficult (Brkan and Bonnet 2020). Gaining access to the rules of an AI system might not be sufficient or desirable to uphold transparency. Consequently, transparency should not limit itself to mere access to the bodywork of an AI system. This has been argued to be insufficient or even not desirable for understanding AI and accountability (Kroll et al. 2016). As explained by the UK House of Lords in its report, transparency might only reveal the technical properties of an AI system, but not necessarily how a decision was made.[44] Therefore, explainability is needed under transparency.[45] Doshi-Velez and Kortz (2017) interpret explanation for a decision as a 'set of abstracted reasons or justifications for a particular outcome, not a description of the decision-making process in general'. Specifically, they require information to be interpretable by humans and to contain the logic utilised in selecting specific inputs and reaching a particular conclusion (Doshi-Velez and Kortz 2017, 4). However, the rights of deploying companies represent legal obstacles to achieving the transparency and, with it, sufficient explainability. The AI Act proposal has failed to address these existing impediments which might be exacerbated, and power/information asymmetries created.

Dignum (2021) warns that these technologies are not to generate trust when isolated, but instead 'trust in AI needs to be derived from trust on the socio-technical system embedding of AI'. In the context of authorities using such systems, this means that institutional arrangements and commitments stand at the forefront of good administrative practices embedding AI systems for the benefit of citizens. Automation can offer many benefits for the enhancement of transparency and accountability of governmental decision-making (Zalnieriute, Moses, and Williams 2019, 16). In terms of human oversight imposed by Article 14 of the AI Act proposal, requiring the public body that relies on AI systems in its decision-making to state how other available information or alternative outcomes were considered in reaching a decision would strengthen this obligation in a way that enables sufficient transparency for impacted individuals (Fink 2021). However, it can also decrease transparency in omitting reasons for a decision made, thereby creating a *de facto* information asymmetry. This is especially the case when, what Burrell (2016) qualified as 'intentional secrecy', is present, because public authorities are bound not to disclose certain information protected either by a contract or a trade secret (Zalnieriute, Moses, and Williams 2019, 17). These aspects are highly problematic and should be addressed by the regulation on AI systems.

Overall, the AI Act proposal is a much-needed steppingstone into the regulation of these new technologies. However, it fails to address the fundamental power imbalances between those who are developing and deploying the technology, and those who are subject to it (Skelton 2021). When observed altogether, the AI Act in its current form focuses too much on technical standards when contrasted with individuals having little in the way of protection or redress. It appears that considerations stop beyond the user – namely, those interacting with individuals who are at the end of the chain (MacCarthy and Propp 2021). The difference between what is said in recitals and the actual text of the regulation make it a regulation for companies, not people. A company seeking to ensure compliance can merely tick the boxes, even though this is insufficient to meet the needs of a public authority using its AI system, since it will not allow them to

protect the fundamental rights of citizens. This in itself can be a threat to compliance with the principle of good administration, as private companies creating these systems have different interests in comparison with public bodies. It is uncertain whether public bodies will be able to transform and fit these self-made standards into their *modus operandi* requiring the preservation of citizens' rights. This should not be the case. Rather, the AI regulatory framework should account for potential power asymmetry and make sure that certain imperative values are imposed *ex ante* through law with and not through soft law language of ethics which has little to no value to EU market competitors.

4. How then to regulate AI?

There are many possibilities to regulate AI – be it through soft law, hard law, or a combination of the two via different types of regulatory systems. They should be carefully considered, as not all of them are suitable.

4.1. Soft law versus *hard law*

The separation of law and morals/ethics is not new. It has been one of the features of utilitarianism for centuries (Hart and Adolphus 1957). In these times, one has to ponder whether transparency is a matter of law or morals.[46] After all, initiators of this idea lived in times where no machines existed, far from implications of AI. Regardless of its nature, transparency tends to be afforded an ethical value and placed within soft law frameworks. Today, the role of soft law has gained more relevance with technological developments outpacing law-making capabilities. The importance of transparency of AI systems in relation to potential consequences identified by the previous section requires careful consideration of its placement within the law.

Soft law consists of non-binding instruments with practical implications. It is characterised by 'substantive expectations which are not directly enforceable' and embodied in instruments such as guidelines, codes of conduct, best practice, public-private partnerships and similar (Marchant 2019). It is often used as a complement to law in a way that it aids national actors in interpreting hard law (Stefan 2012). Perhaps surprisingly, soft law does not necessarily imply less efficacy than rules of hard law (Hage 2018, 38). It has been deemed an important tool in the coping of the governance frameworks with the fast pace of technological developments. Moreover, a plethora of case law from CJEU illustrates that, regardless of its non-legally binding nature, soft law can nevertheless produce binding effects.[47] Its benefits include its quick adoption and revision and thereby adaptation without the need for usual formalities that a typical legal act has to go through (Meyer 2009; Marchant 2019). This also means that soft law is easier to revise in response to developments, which makes it a more attractive choice in the field of technology. Such is the case with AI because it is evolving rapidly. Soft law instruments can thus prove invaluable if they provide technical guidance to those deploying, using and overseeing AI systems in Member States. In the development of AI systems, such measures can include coding machine ethics into AI systems or creating oversight systems.

Soft law as a concept has been subject to criticism as being redundant and detrimental (Klabbers 1998). However, with new challenges for the law, the scale of opposition started

tipping towards reconsidering its role in the regulation of fast-developing fields. For instance, Floridi et al. (2018, 694) emphasise the dual advantage of – what they call – an ethical approach when it comes to AI systems. With an ethical approach, organisations can make use of the social value that AI enables and can anticipate and mitigate costly mistakes. This approach can only function in an ecosystem of public trust where responsibilities are clearer (Floridi et al. 2018, 694). Thus, some form of transparency is needed. To this end, Marchant (2019) proposes complementary approaches to reduce the uncertainties stemming from soft laws in AI. One of them is the establishment of certification authorities who would confirm that a particular set of guidelines has been complied with by an AI system developing company. Alternatively, a coordinating body could bridge standards in making sure that there are no overlaps. However, the existing self-certification as allowed by the AI Act proposal is not the way to go. Codes of conduct can be a strong instrument in ensuring compliance with certain ethical principles. Doctors, lawyers and other impactful professions are prime examples of how this works in practice. Perhaps AI scientists too could be included in this list (Dignum 2019, 99).

Although they are very much welcome as a complement to hard laws, their main danger is this lack of enforceability (Eliantonio and Stefan 2018, 458). Disadvantages of soft law have been a topic of many scholarly discussions. Certain sceptics consider ethical deliberations as a tactic of the industry to buy time, distract the public and delay effective regulation – the so-called 'ethics washing' – thereby creating fake ethics bodies which are short-lived as needed (Metzinger 2019). Another issue of soft law lies in its equally soft and vague language which makes compliance more onerous. It brings a risk of diversification of approaches and standards amongst stakeholders due to its experimental nature. Moreover, the unpredictability of potentially fast-changing soft law undermines legal certainty (Creyke 2010, 18). This risk is unacceptable in any area where individuals' rights can be greatly impacted. Thus, leaving transparency of AI systems to mere codes of conduct and self-regulation by private parties who are motivated by potential profits rather than the impact on the end users, is unacceptable.

This unpredictability in soft law standards also potentially places a burden on public authorities who must comply with different standards than private parties who might not observe the same standards thereby impeding smooth and safe utilisation of AI systems in the public sector for the benefit of citizens and endangering the principle of good governance. Regardless, some authors argue that soft law is used by administrative authorities as a narrowing tool in individual cases when hard law provisions feature ambiguous terms (Tollenaar 2012). However, administrative bodies are expected to follow *inter alia* two distinct principles – legal certainty and proportionate decision-making. The former requires a certain level of regularity in the decision-making behaviour and the latter requires individual interests to be weighed. These principles can collide when soft law is used. Unlike hard law regulation, soft law does not foster trust and reassurance in the public authorities because it lacks legal validity in its creation (Westerman et al. 2018; Senden 2004). When such authorities are using AI systems in their mandatory interaction with citizens and their rights, this trust is pivotal. The AI Act is an important step in this context.

Although some authors argue that the age of hard law is 'gradually dying' and that soft law is becoming the primary tool of governance of new technologies, hard law should be there to preserve interests which would not be considered by self-regulatory soft law

regimes. It is said that 'Soft law is hard law in the making' (Sulev 2020, 1560). While there is some truth to this, some interests can be better protected by public parties. Certain interests are too important to stay at the level of soft law imbued with a lack of proper safeguards. Since soft law lacks oversight and accountability (Hagemann, Skees, and Thierer 2018), there is a risk of abuse by companies which are enabled to create their own 'rules' tacitly without much transparency of the process and thereby isolate certain pivotal considerations which might be too burdensome for them. If we consider ethics as a driving force of laws that eventually embody them, it should come as no surprise that transparency should not feature in soft law instruments such as voluntary codes of conduct.

4.2. Command-and-control versus *self-regulation* versus *co-regulation*

Each industry requires a specific approach to regulation and there are plenty to choose from. Three main modes of regulation are (1) command-and-control, (2) self-regulation, and (3) co-regulation. Command-and-control regulation (1) is characterised by 'the idea of control by a superior' – usually regulatory bodies – supported by sanctions (Ogus 1994, 2; Baldwin, Cave, and Lodge 2011, 106). However, due to their over-burdening nature, they tend to encourage (voluntary) self-regulation (2) of the industry – meaning that the actors are to create rules, monitor and resolve issues amongst themselves before turning to the state (Eijlander 2005). Co-regulation (3) is a complement to legislation and a hybrid of the previous two, which consists of private regulators and the state cooperating in joint institutions (Senden 2005).

The main advantage of command-and-control regulation is that 'the force of law can be used to impose fixed standards with immediacy and to prohibit activity not conforming to such standards' (Baldwin, Cave, and Lodge 2011, 107). Nonetheless, the regulator risks a 'capture' by those regulated, as it has to rely on the information held by them (Mitnick 1980; Quirk 1981). Another concern regarding this type of regulation is the potential for over-regulating by producing unnecessarily complex and rigid rules which are difficult to comply with (Stewart 1988; Teubner 1987). Moreover, it can result in *de facto* exclusion of entrants to the market – as was the case with telecommunications in the UK (OFTEL 1998). Finally, enforcement of such rules is said to be difficult, expensive and with uncertain effects (Baldwin, Cave, and Lodge 2011, 110). A way to counter these disadvantages is by employing incentive-based regimes. However, the uncertainty pertaining to the consequences of incentivising does not fit the field of AI, as it does not solve the problems previously described. Further innovation of AI should be encouraged and incentivised, but not in a way of one-sided regulation, because its nature, complexity and the pace of development require some form of broader, more inclusive participation.

Self-regulation is an alternative to command-and-control regulation. Its advantage rests on the fact that it is efficient and encompasses a higher level of expertise (Baldwin, Cave, and Lodge 2011, 139–140). Since industry actors are in the field and bear the costs themselves, they produce controls efficiently in the form of voluntary codes of conduct and other instruments. Moreover, firms are more inclined to commit to their own rules. Regardless, the costs for the public purse approving these regimes may be considerable, and the self-regulatory rules are not immune from difficulties recognised in the command-and-control systems. In its most basic form, self-regulation is risky. This is especially the case in the digital world. For instance, the difference in data

protection regimes between the US and the EU illustrate this (Hirsch 2011). The procedures in self-regulation may escape the grasp of the public since they can lack transparency, accountability and public approval. Self-regulation thus contains a risk of abuse by companies which are enabled to create their own 'rules' tacitly without much transparency of the process and thereby isolate certain crucial considerations which might be too burdensome for them, yet important for citizens.

Co-regulation involves a process that initially appears like self-regulation, but the state has the final say on drafted soft-law instruments in a way that complies with the relevant law. In this way, private actors participate in the regulation of the industry and the state ensures that drafted regulation meets the statutory requirements (Freeman 1997; Hirsch 2011). Co-regulation enables actors to ensure that the objectives defined by the legislature can be achieved in the context of measures carried out by parties recognised within the field of regulation concerned. It has been favoured as introducing more cost-effective, flexible standards which provide meaningful protection. If unfavourable results are produced, there is still an option to introduce standard legislation. Moreover, unlike self-regulation, the principle of transparency is applied to co-regulation. Thus, if done well, co-regulation can bring the best of both worlds – expertise and adequate protection of interests through law.

The big word 'regulation' tends to strike fear, as it forecasts yet another set of norms to comply with, threatens to stifle innovation and progress, and raises questions about insufficiency of laws to deal with the complexities of AI (Dignum 2019, 97). Regardless, it is important to note that regulation can be beneficial when it is made in a way to incentivise and encourage investments, and further scientific development of AI systems through its norms. In this way, challenged developers can ensure that they develop an AI system which employs techniques that ensure explainability without endangering efficiency. Command-and-control and self-regulation have been criticised are being too simplistic and political rather than taking into account the continuum of different types of legislation. Instead, co-regulation is advocated to be more productive (Prosser 2008). Co-regulation through standardisation based on the New Legislative Framework is at the heart of the AI Act proposal (Ebers 2021). Since technical expertise does not lie with the legislator, co-regulation is an indispensable in the formation of standards applicable to AI systems.

A regulatory regime is not an isolated system featuring only state actors. It is a complex network of a multitude of interrelated and interdependent actors (Black and Murray 2019). Key elements include shared goals, values, and behaviours. This makes regulation equally complex because it ought to reflect and mitigate existing dichotomies by setting rules in appropriate forms – namely, hard law and/or soft law – and modes – command-and-control, self-regulation or co-regulation. Black and Murray (2019) observed how regulation of new technologies occurred throughout the history from radio and telecommunications to the internet. The history of regulation in these areas illustrates that stakeholders (e.g. NGOs, government bodies, companies) have attempted to reassure governments and consumers that they will adhere to ethical principles through the creation of ethical codes and ethics boards. However, this adherence is necessary in effective regulation but insufficient in the absence of specific conditions which rarely exist in a highly competitive market. Thus, if we leave too much to companies developing AI systems, transparency – and thereby our rights – will be at risk.

5. Conclusions

Transparency is an important tool of trust. This article was inspired by a flawed transparency obligations framework for AI systems. The main flaw is the difficulty of individuals challenging the basis of a decision due to an information asymmetry created by AI development companies. This is one of the unaddressed challenges hidden behind the curtain of innovation and its gradual implementation into the traditional problem-ridden embroidery of public law. Transparency is a pivotal notion in this regard.

The long-awaited proposed AI Act represents an important step towards enabling trust in AI. However, improvements are still needed. Although high-risk systems' providers are obliged to comply with transparency obligations, the proposal does not clarify what information would enable 'sufficient transparency' and leaves the AI development companies to regulate through self-assessment, which is highly problematic because it can lead to information and power asymmetries. Moreover, non-high risk AI systems are barely within the scope of regulation and are thus left to self-regulation. Limited transparency obligations for them can be a problem as the regulation does not account for cases where a non-risk system can have high-risk consequences. Consequently, ethics should not substitute hard law.

As discussed, there are many options for regulating AI. Unfortunately, the separation of ethical and legal frameworks has placed the explainability side of transparency into the sphere of self-regulation. Soft-law treatment of AI systems brings a risk of divergent norms and thus legal uncertainty which would make compliance by public authorities with fundamental principles onerous. Although soft-law instruments like codes of conduct can be quite beneficial in the face of speedy technological developments, certain values are too important to be confined to unenforceable frameworks. Therefore, transparency should not be left to self-regulation but should be shaped by the cooperation of private and public actors in a way that ensures proportionate representation of rights and duties.

A transparency framework, as submitted in the AI Act proposal, leaves a risk of abuse by companies because of divergent priorities and interests. Existing impediments might thus be exacerbated. Most importantly, the AI Act (in its current form) fails to address the power imbalance between private parties developing AI systems and public authorities using them. Rather, it creates interdependence between the two which can cause issues for authorities observing principles of good governance towards individuals. The AI regulatory framework should account for this potential power-and-information asymmetry, ensuring that certain imperative values like transparency are imposed *ex ante* through law with and not through soft law language of ethics, which has little to no value to EU market competitors. This imbalance should, and hopefully will, be addressed in the further evolution of the EU regulatory framework for AI systems.

Notes

1. 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 (and encompassing Annexes 1 to 9) [AI Act proposal].
2. Consolidated version of the Treaty on European Union [2012] OJ C326/13.

3. Case C-521/09 P *Elf Aquitaine SA v European Commission* [2011] ECLI:EU:C:2011:620, paras 148-151.
4. See for instance European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).
5. Member States of the EU, 'Declaration of Cooperation on Artificial Intelligence' (10 April 2018) <https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence> accessed 21 May 2021.
6. European Commission, 'Artificial intelligence: Commission kicks off work on marrying cutting-edge technology and ethical standards' (*ec.europa.eu*, Press Release, 9 March 2018) <https://ec.europa.eu/growth/content/artificial-intelligence-commission-kicks-work-marrying-cutting-edge-technology-and-ethical_en> accessed 21 May 2021.
7. See European Commission, 'Liability for emerging digital technologies' SWD (2018) 137 final; European Commission, 'Artificial Intelligence for Europe' COM (2018) 237 final.
8. See G*uerra and others v. Italy* (App no. 14967/89) ECHR 1998-I 64.
9. High-Level Expert Group on Artificial Intelligence, 'Draft Ethics Guidelines for Trustworthy AI' (First draft, 18 December 2018) <https://digital-strategy.ec.europa.eu/en/library/draft-ethics-guidelines-trustworthy-ai> accessed 21 May 2021.
10. European Commission, 'Over 500 comments received on the draft Ethical Guidelines for Trustworthy Artificial Intelligence' (*digital-strategy.ec.europa.eu – News & Views*, February 2019) <https://digital-strategy.ec.europa.eu/en/news/over-500-comments-received-draft-ethical-guidelines-trustworthy-artificial-intelligence> accessed 21 May 2021.
11. High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (2019).
12. European Commission, 'Building Trust in Human-Centric Artificial Intelligence' COM (2019) 168 final.
13. High-Level Expert Group on Artificial Intelligence, 'Policy and investment recommendations for trustworthy Artificial Intelligence' (Working Document for stakeholders' consultation, Brussels, 26 June 2019).
14. Ibid.
15. 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) [2016] OJ L119/1 [GDPR].
16. See Arts. 13(2f) and 15(1) GDPR.
17. Others have contended that this approach is too narrow and should be made more flexible and functional by aiming to balance the interests of both those affected by the decision as well as innovators; See Selbst and Powles (2017).
18. European Commission, 'On Artificial Intelligence – A European approach to excellence and trust' COM (2020) 65 final.
19. European Commission, 'Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence' (Press Release, 21 April 2021) <https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682> accessed 22 May 2021.
20. See Title II AI Act proposal.
21. See Article 5 (1c) (i) AI Act proposal.
22. See Annex III AI Act proposal; European Commission, 'Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence' (Press Release, 21 April 2021) <https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682> accessed 22 May 2021.
23. See Chapter 2 of the Title III of the AI Act proposal.
24. See Article 10 (2–4) AI Act proposal.
25. Article 13 AI Act proposal; NB: Providers comprise companies developing and deploying AI systems.
26. Paras 1 and 2 of Article 6 Artificial Intelligence Act proposal; NB: 'Users' refers to actors which will use the AI systems (e.g. municipalities).

27. See Annex IV AI Act Proposal.
28. Para 2(b) and (c) of Annex IV.
29. Article 14(4) AI Act proposal.
30. See Annex V AI Act proposal.
31. Article 29 AI Act proposal.
32. Case C-340/89 *Vlassopoulou* [1991] ECR I-2357, para 22.
33. Joined cases 240, 242, 261, 262, 268 and 269/82 *Stichting Sigarettenindustrie v. Commission* [1985] ECR I-3831.
34. Article 33 AI Act proposal.
35. Article 52 AI Act proposal.
36. Article 52(1) AI Act proposal.
37. See for instance Article 52 AI Act proposal.
38. This includes a set of directives and regulations on, for instance, machinery, medical devices, vehicles, and equipment; For a full list, see Annex II of the AI Act proposal.
39. 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 [2008] OJ L 218/82.
40. European Commission, 'Conformity assessment' (*ec.europa.eu*) <https://ec.europa.eu/growth/single-market/goods/building-blocks/conformity-assessment_en> accessed 23 May 2021.
41. See Annex I of the 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 [2012] OJ L316/12.
42. 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 (SWD (2021) 84 final) 248.
43. See Advocate General Saugmandsgaard Øe's Opinion in the case C-160/20 Stichting Rookpreventie Jeugd [2021] ECLI:EU:C:2021:618; Article 4 of the Regulation (EC) No 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council and Commission documents [2001] OJ L145/43 specifically allows for a denial of access to information in the interest of protecting commercial interests.
44. Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing, and Able?* (HL 2017–19, 100) paras 95–100.
45. Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing, and Able?* (HL 2017–19, 100) paras 95–100.
46. In the words of Hart: '[…] one of these standards is a moral standard but not all standards are moral'.
47. Case C-322/88 *Grimaldi* [1989] ECR I-4407, para 18; For an overview of all the case law, see Stan 2012.

Acknowledgements

The author was afforded the opportunity to present an early version of this paper at the annual BILETA conference online. The author is grateful for the interest expressed and creative discussions. The author would like to extend her thanks to her PhD supervisors Sofia Ranchordás and Jeanne Mifsud Bonnici, as well as the reviewers whose comments helped shape this work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ida Varošanec http://orcid.org/0000-0001-7271-3091

References

Alì, G. S., and R. Yu. 2021. "Artificial Intelligence Between Transparency and Secrecy: From the EC Whitepaper to the AIA and Beyond." *European Journal of Law and Technology* 12 (3): 1.

Ananny, M., and K. Crawford. 2018. "Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability." *New Media & Society* 20 (3): 973–989.

Baldwin, R., M. Cave, and M. Lodge. 2011. *Understanding Regulation: Theory, Strategy, and Practice*. Oxford: Oxford University Press.

Ball, C. 2009. "What Is Transparency?" *Public Integrity* 11 (4): 293–308. doi:10.2753/PIN1099-9922110400.

Black, J. 2005. "The Emergence of Risk-Based Regulation and the new Public Risk Management in the United Kingdom." *Public law* 512–549.

Black, J. 2010a. "The Role of Risk in Regulatory Processes." In *The Oxford Handbook of Regulation*, edited by Robert Baldwin, Martin Cave, and Martin Lodge, 301–348. Oxford University Press.

Black, J. 2010b. "Risk-based Regulation: Choices, Practices and Lessons Being Learned." In *Risk and Regulatory Policy: Improving the Governance of Risk*, edited by OECD, 185–224. Paris: OECD Publishing.

Black, J., and A. D. Murray. 2019. "Regulating AI and Machine Learning: Setting the Regulatory Agenda." *European Journal of Law and Technology* 10 (3): 1–21.

Brkan, M., and G. Bonnet. 2020. "Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas." *European Journal of Risk Regulation* 11 (1): 18–50.

Buijze, A. 2013. "The Principle of Transparency in EU Law." Diss., Uitgeverij BOXPress.

Burrell, J. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society*, 3 (1): 205395171562251. doi:10.1177/2053951715622512.

Casey, B., A. Farhangi, and R. Vogl. 2019. "Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise." *Berkeley Technology Law Journal* 34: 143–188. doi:10.15779/Z38M32N986.

Creyke, R. 2010. "'Soft Law' and Administrative Law: A New Challenge." *AIAL Forum* 61: 15–22.

Diakopoulos, N. 2016. "Accountability in Algorithmic Decision Making." *Communications of the ACM* 59 (2): 56–62.

Dignum, V. 2019. *Responsible Artificial Intelligence – How to Develop and Use AI in a Responsible Way*. Switzerland: Springer.

Dignum, V. 2021. "On the European AI Act: Acting Is Key." LinkedIn. LinkedIn, June 23, 2021. https://www.linkedin.com/pulse/european-ai-act-acting-key-virginia-dignum/.

Doshi-Velez, F., and M. Kortz. 2017. Accountability of AI Under the Law: The Role of Explanation. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper.

Duval, A. 2021. "Explainable AI, the Key to Open 'Black Boxes.'" Web log. *Towards Data Science* (blog), June 30, 2021. https://towardsdatascience.com/explainable-ai-the-key-to-open-black-boxes-4ad09e04d791.

Ebers, M. 2021. "Standardizing AI-The Case of the European Commission's Proposal for an Artificial Intelligence Act." SSRN. https://ssrn.com/abstract=3900378.

Edwards, L., and M. Veale. 2017. "Slave to the Algorithm: Why a Right to an Explanation is Probably not the Remedy you are Looking for." *Duke L. & Tech. Rev* 16: 18.

Eijlander, P. 2005. "Possibilities and Constraints in the use of Self-Regulation and co-Regulation in Legislative Policy: Experiences in the Netherlands – Lessons to be Learned for the EU?". SSRN. https://ssrn.com/abstract=959148.

Eliantonio, M., and O. Stefan. 2018. "Soft Law Before the European Courts: Discovering a 'Common Pattern'?." *Yearbook of European Law* 37: 457–469. doi:10.1093/yel/yey017.

Eubanks, V. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
Ewald, F. 1991. "Insurance and Risk." In *The Foucault Effect*, edited by Gragam Burchell, Colin Gordon, and Peter Miller, 197–210. Chicago: The University of Chicago Press.
Felzmann, H., E. F. Villaronga, C. Lutz, and A. Tamò-Larrieux. 2019. "Transparency you Can Trust: Transparency Requirements for Artificial Intelligence Between Legal Norms and Contextual Concerns." *Big Data & Society* 6 (1): 2053951719860542.
Fink, M. 2021. "The EU Artificial Intelligence Act and Access to Justice." *EU Law Live*.
Fisher, E. 2010. *Risk Regulation and Administrative Constitutionalism*. Oxford: Hart.
Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, et al. 2018. "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 28 (4): 689–707. doi:10.1007/s11023-018-9482-5.
Freeman, J. 1997. "Collaborative Governance in the Administrative State." *UCLA L. Rev* 45: 1.
Gaumond, E. 2021. Artificial Intelligence Act: What Is the European Approach for AI?" Web log. *Lawfareblog* (blog). https://www.lawfareblog.com/artificial-intelligence-act-what-european-approach-ai.
Ghani, R. 2016. "You Say You Want Transparency and Interpretability?" Web log. *Rayidghani* (blog) http://www.rayidghani.com/2016/04/29/you-say-you-want-transparency-and-interpretability/.
Goodman, B., and S. Flaxman. 2017. "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation.'" *AI Magazine* 38 (3): 50–57. doi:10.1609/aimag.v38i3.2741.
Hage, J. 2018. "What Is Legal Validity? Lessons from Soft law." In *Legal Validity and Soft Law*, edited by P. Westerman, J. Hage, S. Kirste, and A Mackor, 19–45. Cham: Springer.
Hagemann, R., J. H. Skees, and A. Thierer. 2018. "Soft law for Hard Problems: The Governance of Emerging Technologies in an Uncertain Future." *Colo. Tech. LJ* 17: 37.
Hart, H., and L. Adolphus. 1957. "Positivism and the Separation of Law and Morals." *Harvard Law Review* 71: 593.
Heald, D. 2012. "Transparency as an Instrumental Value." In *Transparency: The Key to Better Governance?*, edited by Christopher Hood and David Heald, 59–73. Oxford: British Academy. British Academy Scholarship Online.
Heald, D. 2012. "Varieties of Transparency." In *Transparency: The Key to Better Governance?*, edited by Christopher Hood, and David Heald, 25–43. Oxford: British Academy. British Academy Scholarship Online.
Hildebrandt, M., and S. Gutwirth. 2008. *Profiling the European Citizen: Cross-Disciplinary Perspectives*. Dordrecht, Netherlands: Springer.
Hirsch, D. 2011. "The Law and Policy of Online Privacy: Regulation, Self-Regulation, or Co-Regulation." *Seattle University Law Review* 34: 439.
Kaminski, M. E. 2019. "The Right to Explanation, Explained." *Berkeley Technology Law Journal* 34: 189.
Kingsman, N., E. Kazim, M. Chaudhry, A. Hilliard, A. Koshiyama, R. Polle, G. Pavey, and U. Mohammed. 2021. "Public Sector AI Transparency Standard." *Available at SSRN 3986213*.
Klabbers, J. 1998. "The Undesirability of Soft law." *Nordic Journal of International Law* 67: 381–391.
Kroll, J., J. Huey, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, and H. Yu. 2016. "Accountable Algorithms." *University of Pennsylvania Law Review* 165: 633.
Liu, B. 2021. "In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human–Ai Interaction." *Journal of Computer-Mediated Communication* 26 (6): 384–402. doi:10.1093/jcmc/zmab013.
MacCarthy, M., and K. Propp. 2021. "Machines Learn that Brussels Writes the Rules: The EU's New AI Regulation." *Brookings*.
Marchant, G. 2019. "'Soft Law' Governance of Artificial Intelligence." https://aipulse.org/soft-law-governance-of-artificial-intelligence/.
Metzinger, T. 2019. "Ethics Washing Made in Europe." Der Tagesspiegel, April 8, 2019. https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html.
Meyer, Timothy. 2009. "Soft Law as Delegation." *Fordham International Law Journal* 32 (3): 888.
Mitnick, B. M. 1980. *The Political Economy of Regulation*. New York: Columbia University Press.

OFTEL. 1998. *Second Submission to the Culture, Media and Sport Select Committee: Beyond the Telephone, the Television and the PC – Regulation of the Electronic Communications Industry*. London.

Ogus, A. 1994. *Regulation: Legal Form and Economic Theory*. Oxford: Clarendon Press.

Opdebeek, I., and S. De Somer. 2016. "The Duty to Give Reasons in the European Legal Area: A Mechanism for Transparent and Accountable Administrative Decision-Making? A Comparison of Belgian, Dutch, French and EU Administrative Law." *Rocznik Administracji Publicznej* 2: 97–148.

Pasquale, F. 2016. *The Black Box Society: The Secret Algorithms Behind Money and Information*. Cambridge, MA: Harvard University Press.

Prosser, T. 2008. "Self-regulation, co-Regulation and the Audio-Visual Media Services Directive." *Journal of Consumer Policy* 31 (1): 99–113.

Quirk, P. J. 1981. *Industry Influence in Federal Regulatory Agencies*. Princeton, NJ: Princeton University Press.

Reed, C., K. Grieman, and J. Early. 2021. "Non-Asimov Explanations Regulating AI through Transparency." SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3970518.

Reichel, J. 2021. "Openness and Transparency." In *The Oxford Handbook of Comparative Administrative Law*, edited by Peter Cane, Herwig Hofmann, Ip Eric Chi Yeung, and Peter L. Lindseth, 935–956. Oxford: Oxford University Press.

Rouvroy, A., and Y. Poullet. 2009. "The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy for Democracy." Essay. In *Reinventing Data Protection?*, edited by Serge Gutwirth, Yyves Poullet, Paul De Hert, Cécile De Terwangne, and Sjaak Nouwt, 45–76. Dordrecht: Springer.

Selbst, A. D., and J. Powles. 2017. "Meaningful Information and the Right to Explanation." *International Data Privacy Law* 7 (4): 233–242. doi:10.1093/idpl/ipx022.

Senden, L. 2004. *Soft Law in European Community Law*. Vol 1. Oxford: Hart Publishing.

Senden, L. 2005. "Soft law, Self-Regulation and co-Regulation in European law: Where do They Meet?" *Electronic Journal of Comparative Law* 9 (1): 1–27.

Skelton, S. K. 2021. "Europe's Proposed AI Regulation Falls Short on Protecting Rights." *ComputerWeekly.com*. https://www.computerweekly.com/feature/Europes-proposed-AI-regulation-falls-short-on-protecting-rights.

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

Stefan, O. 2012. "European Union Soft Law: New Developments Concerning the Divide Between Legally Binding Force and Legal Effects." *The Modern Law Review* 75 (5): 879–893.

Stewart, R. B. 1988. "Regulation and the Crisis of Legalisation in the United States." In *Law as an Instrument of Economic Policy – Comparative and Critical Approaches*, edited by Terence Daintith, 97–134. Berlin: De Gruyter.

Stirling, A. 1998. "Risk at a Turning Point?" *Journal of Risk Research* 1 (2): 97–109. doi:10.1080/136698798377204.

Stohl, C., M. Stohl, and P. M. Leonardi. 2016. "Digital age| Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital age." *International Journal of Communication* 10: 15.

Sulev, G. 2020. "Soft Law in EU Electronic Communications Regulation: A Bulgarian Case Stud ." *European Papers* 5 (3): 1555–1564.

Teubner, G. 1987. *Juridification of Social Spheres: A Comparative Analysis in the Areas of Labor, Corporate, Antitrust and Social Welfare Law*. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110921472.

Tollenaar, A. 2012. "Soft Law and Policy Rules in the Netherlands." SSRN. https://ssrn.com/abstract=1904427

Tridimas, T. 2006. *The General Principles of EU Law*. Oxford: Oxford University Press.

Van Dijk, N., R. Gellert, and K. Rommetveit. 2016. "A Risk to a Right? Beyond Data Protection Risk Assessments." *Computer Law & Security Review* 32 (2): 286–306.

Veale, M., and F. Z. Borgesius. 2021. "Demystifying the Draft EU Artificial Intelligence Act – Analysing the Good, the bad, and the Unclear Elements of the Proposed Approach." *Computer Law Review International* 22 (4): 97–112.

Volpato, A. 2021. "'Part of Eu Law', but Only Partially: The Issue of the Accessibility of Harmonised Standards." Web log. *Maastrichtuniversity.nl* (blog), October 15, 2021. https://www.maastrichtuniversity.nl/blog/2021/10/%E2%80%9Cpart-eu-law%E2%80%9D-only-partially-issue-accessibility-harmonised-standards.

von Eschenbach, W. J. 2021. "Transparency and the Black box Problem: Why we do not Trust AI." *Philosophy & Technology* 34 (4): 1607–1622.

Wachter, S., B. Mittelstadt, and L. Floridi. 2017. "Why a Right to Explanation of Automated Decision-Making Does not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7 (2): 76–99.

Wade, W., and C. F. Forsyth. 2000. *Administrative Law*. 8th ed. Oxford: Oxford University Press.

Weller, A. 2019. "Challenges for Transparency." https://arxiv.org/pdf/1708.01870.

Westerman, P., J. Hage, S. Kirste, and A. R. Mackor, eds. 2018. *Legal Validity and Soft Law. Vol. 122*, 1–266. Cham: Springer

Wischmeyer, T. 2020. "Artificial Intelligence and Transparency: Opening the Black box." In *Regulating Artificial Intelligence*, edited by Thomas Wischmeyer, and Timo Rademacher. Cham: Springer.

Wynne, B. 2002. "Risk and Environment as Legitimatory Discourses of Technology: Reflexivity Inside out?" *Current Sociology* 50 (3): 459–477. doi:10.1177/0011392102050003010.

Zalnieriute, M., L. B. Moses, and G. Williams. 2019. "The Rule of law and Automation of Government Decision-Making." *The Modern Law Review* 82 (3): 425–455.

Zerilli, J., A. Knott, J. Maclaurin, and C. Gavaghan. 2019. "Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?" *Philosophy & Technology* 32 (4): 661–683.