INTERNET POLICY REVIEW
Journal on internet regulation

Volume 10 **Issue** 3

RESEARCH ARTICLE

OPEN ACCESS

PEER REVIEWED

Beyond the individual: governing AI's societal harm

Nathalie A. Smuha *KU Leuven* nathalie.smuha@gmail.com

DOI: https://doi.org/10.14763/2021.3.1574

Published: 30 September 2021
Received: 14 December 2020 **Accepted:** 19 March 2021

Funding: Research Foundation Flanders (FWO)
Competing Interests: The author has declared that no competing interests exist that have influenced the text.

Citation: Smuha, N. A. (2021). Beyond the individual: governing AI's societal harm. *Internet Policy Review*, *10*(3). https://doi.org/10.14763/2021.3.1574

Keywords: Artificial intelligence, Society, Environmental law, EU law, AI governance

Abstract: In this paper, I distinguish three types of harm that can arise in the context of artificial intelligence (AI): individual harm, collective harm and societal harm. Societal harm is often overlooked, yet not reducible to the two former types of harm. Moreover, mechanisms to tackle individual and collective harm raised by AI are not always suitable to counter societal harm. As a result, policymakers' gap analysis of the current legal framework for AI not only risks being incomplete, but proposals for new legislation to bridge these gaps may also inadequately protect societal interests that are adversely impacted by AI. By conceptualising AI's societal harm, I argue that a shift in perspective is needed beyond the individual, towards a regulatory approach of AI that addresses its effects on society at large. Drawing on a legal domain specifically aimed at protecting a societal interest—environmental law—I identify three 'societal' mechanisms that EU policymakers should consider in the context of AI. These concern (1) public oversight mechanisms to increase accountability, including mandatory impact assessments with the opportunity to provide societal feedback; (2) public monitoring mechanisms to ensure independent information gathering and dissemination about AI's societal impact; and (3) the introduction of procedural rights with a societal dimension, including a right to access to information, access to justice, and participation in public decision-making on AI, regardless of the demonstration of individual harm. Finally, I consider to what extent the European Commission's new proposal for an AI regulation takes these mechanisms into consideration, before offering concluding remarks.

This paper is part of **Governing "European values" inside data flows**, a special issue of *Internet Policy Review* guest-edited by Kristina Irion, Mira Burri, Ans Kolk, Stefania Milan.

Introduction

Artificial Intelligence (AI), an umbrella term for a range of technologies that are considered to demonstrate 'intelligent' behaviour, plays an increasingly important role in all domains of our lives. A distinction is often made between *reasoning-based* or *code-driven AI* on the one hand, and *data-driven or learning-based AI* on the other hand (High-Level Expert Group on AI, 2019a). The former covers techniques that rely primarily on the codification of symbols and rules, based on which the system 'reasons' (using a top-down approach to design the system's behaviour), whereas the latter covers techniques that rely primarily on large amounts of data, based on which the system 'learns' (using a bottom-up approach to design the system's behaviour) (Hildebrandt, 2018; Dignum, 2019). The distinction between both should however not be seen as strict; models can be hybrid and incorporate elements of both techniques. In this paper, the focus lays on data-driven AI systems, given their reliance on data flows.

Spurred by increased computing capacity and data availability, AI systems can be deployed in a manner that generates significant benefits to individuals, groups and society at large. At the same time, they also raise the possibility of significant harm, for instance by breaching fundamental rights or causing other adverse effects (O'Neil, 2017; Russell, 2019; Yeung, 2019b; Solow-Niederman, 2020; Muller, 2020; Gebru, 2020).

The question can be asked to which extent the challenges raised by AI are new or a mere reiteration of the challenges raised by other new technologies. While certainly not unique in terms of the risks they entail, as argued elsewhere, the most promising features of AI applications are also liable to exacerbate these risks (Smuha, 2021). These features include, amongst others, AI systems' self-learning ability and hence potential unpredictability, the vast scale and speed at which they can operate, their ability to distil information from data that may escape the human eye, the opportunity they provide to delegate human authority and control over the execution of—sometimes highly sensitive—decisions and tasks, as well as their opaque decision-making processes, which render it difficult to assess and evaluate compliance with legislation (Mittelstadt et al., 2016; O'Neil, 2017; Solow-Niederman, 2019).

An increasing number of actors—including researchers, journalists, private companies, public entities and civil society organisations—are trying to map how the development and use of AI can cause harm, for instance by breaching fundamental rights or causing other adverse effects (Zuiderveen Borgesius, 2018; Crawford et al., 2019; Yeung et al., 2020; CAHAI, 2020; European Union Agency for Fundamental Rights, 2020; Hao, 2021). Furthermore, across the world, regulators are starting to assess the extent to which existing laws are able to counter these harms, or whether new regulatory measures may be needed to secure protection therefrom (Council of Europe, 2019; UNESCO, 2020). The European Commission, for instance, is currently conducting such an assessment as regards the EU legal order (European Commission, 2020a; 2020c), and recently proposed a new AI-specific regulation (European Commission, 2021).

The legal assessments currently undertaken, and the risk analysis related thereto, are focusing primarily on AI's adverse impact on individuals and—to a lesser extent—on specific groups or collectives of individuals (Mittelstadt, 2017; Taylor et al., 2017). The use of AI-systems can, however, also cause *societal harm*, which can be distinguished from—and which can transcend—*individual harm* and *collective harm*. The fact that certain uses of AI-systems for instance risk harming the democratic process, eroding the rule of law or exacerbating inequality goes beyond the concern of (the sum of) individuals but affects society at large (Bayamlıoğlu & Leenes, 2018; Zuiderveen Borgesius et al., 2018; Brkan, 2019).

While the societal impact of AI systems is increasingly discussed—particularly under the influence of STS studies—AI-enabled societal harm has so far been less examined from a legal perspective. Such examination is more difficult to conduct, as the contours of AI's impact on societal interests are less tangible and hence more difficult to conceptualise in legal terms (Van der Sloot, 2017; Yeung, 2019a). As a consequence, policymakers risk making an incomplete analysis of the legal gaps they should tackle to secure comprehensive protection against AI's adverse effects. In addition, by overlooking this societal dimension, the legal measures they may propose to address gaps in the legal framework can likewise prove inadequate.

This risk is particularly salient given the predominantly individualistic focus of the current legal system, such as data protection law, but also procedural law more generally (van der Sloot & van Schendel, 2021). Indeed, the manner in which EU law currently addresses AI-related harm primarily hinges on private enforcement, by relying on individuals to challenge potentially harmful practices. These challenges can in principle only be initiated by individuals able to establish the infringement of an individual right—such as the right to data protection or non-dis-

crimination—or another directly suffered demonstrable harm to their private interests. Societal harm is however not always reducible to instances of individual harm (Kutz, 2000). Moreover, in some cases, even when both types of harm do overlap, individual harm may be negligible or indiscernible, and hence an insufficient ground to challenge the harmful practice. The following conundrum hence arises: how can we reconcile the need to protect societal interests adversely impacted by AI in the context of a legal system that primarily focuses on individual rights and remedies?

This conundrum is not unique to the AI-context. An analogy can, for example, be drawn with environmental harm, which likewise encompasses a societal dimension that cannot always be reduced to demonstrable individual harm. This resulted in the creation of new legal mechanisms to safeguard environmental interests at the EU level (van Calster & Reins, 2017). Accordingly, to secure protection against AI's societal harms, a shift from an individualistic approach towards one that also embodies a societal perspective is warranted. This shift first requires a legal conceptualisation of AI's societal harms, based on which gaps in protection can be identified and addressed. The importance of considering the societal adverse impact of the use of AI and other data-driven technologies was already stressed by other scholars (Hildebrandt, 2018; Yeung, 2019a; Cohen, 2019; Véliz, 2020; Viljoen, 2020; van der Sloot & van Schendel, 2021). In this paper, I build thereon with the aim of clarifying the legal protection gap in EU law and identifying mechanisms that EU policymakers can consider when tackling this issue.

To this end, I start by distinguishing the three above mentioned types of harm that AI systems can generate (2). Although the societal harms raised by AI can be very diverse, I identify some common features through which they can be conceptualised (3). Next, I venture into a parallel with a legal domain specifically aimed at protecting such an interest: EU environmental law (4). Based on the legal mechanisms adopted under environmental law with a distinct societal dimension, I draw a number of lessons for EU policymaking in the context of AI (5). Finally, I briefly evaluate the European Commission's proposed AI regulation in light of those lessons (6), before providing concluding remarks (7).

2. Individual, collective and societal harm

For the purpose of this paper, harm is conceptualised as a wrongful setback to or thwarting of an *interest* (Feinberg, 1984), under which I also include harm in the non-physical sense, such as the breach of a right. [1] On this basis, I distinguish three types of interests—and hence three types of harm—that should be consid-

ered in the context of AI governance: individual harm, collective harm and societal harm. These harms are connected to each other, yet they can also be assessed in their own right. Evidently, the use of AI systems can also generate individual, collective and societal benefits, by positively impacting these respective underlying interests. In this paper, however, I focus on AI's potential harms rather than its benefits. As the terms individual harm, collective harm and societal harm have been used in different ways by different authors, in what follows I provide a description of my understanding of each.

Individual harm occurs when one or more interests of an individual are wrongfully thwarted. [2] This is the case, for instance, when the use of a biased facial recognition system—whether in the context of law enforcement or in other domains—leads to wrongful discrimination against people of colour. Of course, the thwarting of such interest does not occur in isolation from a social, historical and political context (Winner, 1980; Simon, 1995)—after all, AI systems are socio-technical systems (Hasselbalch, 2019; High-Level Expert Group on AI, 2019b; Theodorou & Dignum, 2020; Ala-Pietilä & Smuha, 2021). Nevertheless, in this scenario, at the receiving end of the harm stands an identifiable individual.

Collective harm occurs when one or more interests of a collective or group of individuals are wrongfully thwarted. Just as a collective consists of the sum of individuals, so does this harm consist of the sum of harms suffered by individual members of the collective. The use of the abovementioned biased facial recognition system, for instance, can give rise to collective harm, in so far as it thwarts the interest of a specific collective of people—namely people of colour who are subjected to the AI system—not to be discriminated against. The collective dimension thus arises from the accumulation of similarly thwarted individual interests. The harmed individuals can be complete strangers to each other (like in the above example, where only their skin colour connects them) or they can be part of an (in)formal group.

Societal harm occurs when one or more interests of society are wrongfully thwarted. In contrast with the above, societal harm is thus not concerned with the interests of a particular individual or the interests shared by a collective of individuals. Instead, it concerns harm to an interest held by society at large, going over and above the sum of individual interests. This can be illustrated with a few examples.

1. Some, but not all, wrongful setbacks to interests are protected by law. It should also be noted that the concept of harm is not static. It changes over time, along with the normative framework of a given society. See in this regard Conaghan, 2002.
2. These interests can, for instance, be physical, psychological, financial, social or legal in nature.

3. AI and societal harm

3.1 Three examples: impact on equality, democracy and the rule of law

Let's start by revisiting the above example of the facial recognition system. First, by making use of such a biased system and wrongfully thwarting the interest of an individual of colour, the system's deployer can cause individual harm. The accumulation of the harm done to individuals of colour at the collective level, entails collective harm. Yet a third type of harm is at play. Whether individuals are coloured or not, and whether they are subjected to the particular AI system or not, they share a higher interest to live in a society that does not discriminate against people based on their skin colour and that treats its citizens equally. That interest is different from the interest not to be discriminated against, and can hence be distinguished from the individual or collective harm done to those directly subjected to the AI system. In other words, societal harm may well include instances of individual and collective harm, but has an impact beyond it. It can hence be assessed as a *sui generis* type of harm. [3]

Besides the interest of equality, the deployment of AI systems can adversely impact a range of other societal interests. Consider this second example. AI systems can be used to collect and analyse personal data for profiling purposes, and subsequently subject individuals to targeted manipulation (Zuiderveen Borgesius et al., 2018; Brkan, 2019). Scandals like Facebook/Cambridge Analytica made it painfully clear that psychographic targeting can be deployed to shape political opinions with the aim of influencing election outcomes (Isaak & Hanna, 2018). Since individuals—in their capacity of product or service users—continue to share ever more data about themselves in different contexts, data flows steadily increase. Accordingly, these manipulative practices can occur at an ever-wider scale and can yield ever more effective results. The potential harm that can ensue—whether it is election interference, hate-mongering, or societal polarisation—is not limited to the individual who is directly manipulated, but indirectly affects the interests of society at large.

Of course, this example does not occur within a legal vacuum. While in the example above the right to non-discrimination might offer a certain level of solace, in this example some recourse can be found in data protection laws. However, as evoked above, these rights are primarily focused on preventing individual harm. In

3. See in this regard also Emile Durkheim's conceptualisation of society as a *sui generis* entity (Durkheim, 1925).

the case of the EU General Data Protection Regulation (GDPR), individuals are given 'control' of their personal data by equipping them with a mix of rights, including for instance the right to access their personal data and obtain information on the processing thereof, the right to have their data rectified and erased, and the right not to be subject to a decision based solely on automated processing. Their personal data can only be processed in case an appropriate legal basis exists, such as their consent for instance (Cate & Mayer-Schonberger, 2013; Van Hoboken, 2019; Tamo-Larrieux et al., 2020). And while the GDPR not only enshrines individual rights but also imposes certain obligations directly upon data processors and controllers, individuals can, to a certain extent, 'waive' such protection through their consent.

This raises two issues. First, the presence of a proper legal basis for data gathering and analysis is often disputed – even when it concerns consent. Despite legal obligations to this end, privacy notices can be anything but reader-friendly or effective, and may leave individuals unaware of what precisely they are consenting to—and which of the endless list of third-party vendors can use their data (Schermer et al., 2014; Van Alsenoy et al., 2014; Zuiderveen Borgesius, 2015; Barrett, 2018; Bietti, 2020). Second, even assuming that the individual carefully reads, understands, and consents to the use of her data, this still leaves uninvolved and unprotected all those who may be indirectly harmed by the subsequent practices enabled by that data, and hence leaves unaddressed the potential societal harm—such as the breach of integrity of the democratic process. As van der Sloot and van Schendel (2021) state: a legal regime that addresses incidental data harms only on an individual level runs the risk of leaving unaddressed the underlying causes, allowing structural problems to persist. Furthermore, besides harm at the societal level, it can also engender new types of indirect individual or collective harms. In this regard, Viljoen (2020) rightly emphasises that an overly individualistic view of the problem fails to acknowledge data's relationality, and the way in which data obtained from individual A can subsequently be used to target individual B even without having obtained similar data from the latter.

The problem, however, still goes further. Consider a third example. AI systems can be used in the context of law enforcement, public administration or the judicial system, so as to assist public officials in their decision-making processes (Kim et al., 2014; Liu et al., 2019; Zalnieriute et al., 2019; AlgorithmWatch, 2020). These systems embody norms that are guided by the legal rules applicable in a specific situation (e.g., the allocation of welfare benefits, the detection of tax fraud, or the assessment of the risk of recidivism). The outcome of these decisions, which are li-

able to significantly affect legal subjects, hence depend on the way in which legal rules are embedded in the algorithmic system (Hildebrandt, 2015; Binns, 2018). In the case of learning-based AI systems, this process hinges on the system's design choices and the data it is being fed—choices that often remain invisible. Moreover, the internal decision-making process of learning-based systems is often non-transparent and non-explainable, which also renders the explainability and the contestability of the decisions more difficult (Pasquale, 2015; Ananny & Crawford, 2016; Bayamlıoğlu, 2018).

In addition, the public officials in charge of taking the final decision and accountable for its legality, may lack the knowledge to assess the justifiability of the AI system's outcome (Brownsword, 2016; Hildebrandt, 2018; Bayamlıoğlu & Leenes, 2018). This risks diminishing the accountability of public decision-makers for the legality of such decisions, and thereby undermining the broader societal interest of the rule of law (Binns, 2018; Zalnieriute et al., 2019; Buchholtz, 2020). The fact that, today, the outcomes generated by public AI systems are often merely informative or suggestive rather than decisive, is but a cold comfort amidst the existence of substantial backlogs which reduce a thorough review of the decision by public officials to a mere source of delay.

In this example too, different types of harm can be distinguished, which are not entirely reducible to each other. An individual subjected to an AI-informed decision taken by a public actor can suffer individual harm for a range of reasons. The correlations drawn by the AI system can be inapplicable, biased or incorrect, but it is also possible that the legal rule applicable to the situation was erroneously embedded in the system. Assuming the individual's awareness of the problem and depending on the context, she may be able to invoke a right to challenge the AI-informed decision—such as the right to a good administration, the right to a fair trial, or the right to privacy (like in the Dutch SyRIcase [4] for instance). Yet the above-described impact on the rule of law also leads to societal harm which is firmly connected to, yet different from, potential individual harm. It also affects all those who are not directly interacting with or subjected to the decision-making process of the specific AI system, and hence goes over and beyond any (accumulative) individual harm.

In sum, an overly individualistic focus on the harm raised by AI risks overlooking its societal dimension, which should equally be tackled. Importantly, the above

4. *NJCM et. al. and FNV v Staat der Nederlanden*, Rechtbank Den Haag, 5 February 2020, C-09-550982-HA ZA 18-388, ECLI:NL:RBDHA:2020:865.

should not be read as an affirmation that all individual and collective harms raised by AI can already be tackled by mere reliance on individual rights. Also with regard to these harms, legal gaps in protection exist and merit being addressed (European Commission, 2020w; CAHAI, 2020; Smuha, 2020). In this article, however, the focus lays on legal protection against AI's societal harms.

3.2 AI's societal harm: common features and concerns

The three examples referred to above—AI-based facial recognition, AI-based voter manipulation, and AI-based public decision-making—concern three different AI applications impacting (at least) three different societal interests: equality, democracy and the rule of law. It is hence not possible, nor desirable, to reduce AI's potential for societal harm to a monolithic concern. Nevertheless, an examination of the issues raised by such harm reveals some commonalities that are useful for the purpose of a legal conceptualisation.

First, in each case, a particular individual harm occurs and is typically safeguarded by an accompanying individual remedy for protection against such harm. In the first example, an individual's right to non-discrimination is at play; in the second, an individual's right to data protection; in the third, an individual's right to good administration or a fair trial. Yet the potential breach of the right can often only be invoked by the individual concerned, not by a third party. [5] More importantly, the individual harm will often remain unnoticed given the opacity of the way in which AI systems are designed and operate. This opacity is often also accompanied by a lack of transparency of how AI systems are used by the product or service provider. As a consequence, it is not only difficult to be aware of the harm, but it may be even more difficult to demonstrate it and establish a causal link. I call this *the knowledge gap problem*. Moreover, even if there is awareness, the individual harm may be perceived as insignificant, or in any case as too small in proportion to the costs that may be involved when challenging it. Hence, the individual is unlikely prompted to challenge the problematic practice. I call this *the threshold problem*. Furthermore, as noted above, an individual can also consent to the practice or otherwise acquiesce, hence seemingly waiving the opportunity to invoke a protective

5. Article 80 of the General Data Protection Regulation, however, establishes a right for data subjects to mandate certain organisations to lodge a complaint with the national supervisory authority on their behalf. Furthermore, it also allows member states to provide that these organisations—independently of a data subject's mandate—have the right to lodge a complaint with the authority if they consider that the data subject's rights under the regulation were infringed. Hence, under this scenario, a third (specifically qualified) party could in fact undertake independent action. Nevertheless, this third party must still demonstrate that a data subject's right was infringed. Evidence of individual harm thus remains a requirement, which complicates matters in case of potential consent or lack of knowledge.

right in exchange for perceived personal gains—thereby also undermining actions by third parties to invoke this right on their behalf. I call this *the egocentrism problem*.

Second, in each case, in addition to individual harm, there is also an instance of societal harm, as adverse effects occur not only to the individuals directly subjected to the AI system, but to society at large too. In the third example, the integrity of the justice system is an interest shared not only by individuals appearing before a court, but also to those who never set a foot therein. This societal harm concerns a different interest than the individual harm. The harm suffered by an individual who faces discrimination, manipulation or an unjust judicial or administrative decision, is different than the societal harm of the unequal treatment of citizens, electoral interference, or erosion of the rule of law. Nevertheless, the current legal framework primarily focuses on legal remedies for the individuals directly subjected to the practice, rather than to 'society'.

Third, differently than individual or collective harm, societal harm will often manifest itself at a subsequent stage only, namely over the longer term rather than in the immediate period following the AI system's use. This gap in time—especially when combined with the opacity of the AI systems functioning and use—not only complicates the identification of the harm itself, but also of the causal link between the harm and the AI-related practice. An additional obstacle in this regard concerns the 'virtual' nature of the harm. In contrast with, for instance, the harm caused by a massive oil leak or burning forest, the societal harm that can arise from the use of certain AI applications is not as tangible, visible, measurable or predictable—which renders the demonstration of both individual harm and of collective harm challenging.

Fourth, societal harm typically does not arise from a single occurrence of the problematic AI practice. Instead, it is often the widespread, repetitive or accumulative character of the practice that can render it harmful from a societal perspective (Kernohan, 1993). This raises further obstacles. Thus, it may be challenging to instil a sense of responsibility in those who instigate a certain practice, if it is only the accumulative effect of their action together with the action of other actors that causes the harm (Kutz, 2000). Moreover, this phenomenon also gives rise to the difficulty of the *many hands*-problem (Thompson, 1980; van de Poel et al., 2012; Yeung, 2019a). Besides the accumulative effects raised by a certain type of conduct, also the opacity of different types of (interacting) conduct by many hands can contribute to the harm and to the difficulty of identifying and addressing it (Nissenbaum, 1996).

In the context of the wide-spread use of AI systems, not one but three levels of this problem come to mind. First, at the level of the AI system, multiple components developed and operated by multiple actors can interact with each other and cause harm, without it being clear which component or interaction is the direct contributor thereof. Second, at the level of the organisation or institution deploying the system (whether in the public or private sector), different individuals may contribute to a process in many different ways, whereby the resulting practice can cause harm. Last, this issue manifests itself at the level of the network of organisations and institutions that deploy the problematic AI application. The scale of these networks and their potential interconnectivity and interplay renders the identification of the problematic cause virtually impossible—even more so if there isn't necessarily *one* problematic cause. A broadened perspective of AI-enabled harm is hence not only required at the suffering side, but also at the causing side of the harm.

Finally, as was already alluded to above, the societal harm that can arise from the mentioned AI applications is not easily expressible in current *human rights* discourse (Yeung, 2019a). While a particular human right may well be impacted, a one-on-one relationship between such right and the societal harm in question can be lacking. This is related to the fact that many human rights embody an individualistic perspective of harm. As Karen Yeung however clarifies, what is at stake is the destabilisation "*of the social and moral foundations for flourishing democratic societies*" which enable the protection of human rights and freedoms in the first place (Yeung, 2019a). Hence, our intuitive recourse to the legal remedies provided by human rights law in the context of AI's harms will only offer partial solace. Furthermore, even this partial solace is on shaky grounds, in light of what I denoted above as the *knowledge gap problem*, *the threshold problem*, and the *egocentrism problem*.

By no means does this imply that human rights have become obsolete in this field—quite the contrary. As argued elsewhere, an overly individualistic interpretation of human rights overlooks the fact that most human rights also have a clear societal dimension, not least because their protection can be considered a societal good (Smuha, 2020). Moreover, in some cases, individual human rights have been successfully invoked to tackle societal challenges, such as for instance in the 2019 Dutch *Urgenda* case [6]. However, since in these cases a demonstration of individual harm is typically still required, this will not be a uniformly accepted or comprehen-

6. *Staat der Nederlanden v Stichting Urgenda*, Hoge Raad, 20 December 2019, 19/00135, ECLI:NL:HR:2019:2006.

sive solution. Accordingly, while human rights remain both an essential normative framework and important legal safeguard against AI-enabled harm, relying on their private enforcement may be insufficient (van der Sloot & van Schendel, 2021). Policymakers assessing AI's risks should hence broaden their perspectives as concerns both the analysis of harms and the analysis of legal gaps and remedies. While this shift is starting to take place as regards privacy—which an increasing number of scholars convincingly argued should be conceptualised in broader terms, as not just an individual but also a societal good (Lane et al., 2014; Cohen, 2019; Véliz, 2020; Bamberger & Mayse, 2020; Viljoen, 2020)—it must be applied to the wider range of societal interests that can be adversely impacted by the use of AI.

This broadened analysis is not only needed to better understand AI's impact on society. Instead, policymakers should also draw thereon when assessing the legal gaps in the current legislative framework and identifying measures to tackle those gaps. Given its specific features, addressing societal harm may require different measures than addressing individual or collective harm. Legal remedies that hinge solely on (collectives of) individuals who may or may not be able to challenge potential right infringements will not always provide sufficient protection when societal interests are at stake, hence leading to a legal protection gap. Countering societal harm will also require 'societal' means of intervention to safeguard the underlying societal infrastructure enabling human rights, democracy and the rule of law. Given the EU's intentions to address some of the adverse effects generated by AI, in what follows, I take a closer look at what role EU law can play in tackling this legal protection gap.

4. Countering societal harm: environmental law as a case study

To answer the above question, I argue that inspiration can be drawn from a legal domain that is specifically aimed at protecting a societal interest: (EU) environmental law. While a polluting practice or activity —whether undertaken by a public or private actor—is liable to cause individual harm, the interest to secure a clean and healthy environment is one that is shared by society at large. An individual living in city A might be unaware of, choose to ignore, or be indifferent to the polluting practice. Yet the adverse effects that will ensue from the practice are likely to go over and above the individual or collective level, as it can give rise to harm also for people living in city B, country C and region D, and to future generations.

Since a number of similarities exist between environmental harm and the societal

harms described above [7], the solutions provided by EU environmental law could be explored by analogy. Indeed, just as in the examples of societal harm raised by the use of AI, a period of time may lapse between a polluting practice and the tangible manifestation of the environmental harm. Moreover, in many cases, a single instance of the polluting practice can be relatively benign, yet it may be the accumulative or systemic nature of the practice that causes the environmental harm at the societal level. In addition, access to an effective legal remedy by individuals may be challenging when relying on individual rights, especially when individual harm is insignificant, difficult to demonstrate or unknown.

Early on, however, environmental harm has been recognised as affecting a societal interest rather than a (mere) individual one. In other words: since the protection of the environment is important for society at large, it should not solely hinge upon the ability or willingness of individuals to go to court. As a consequence, mechanisms have gradually been adopted in this area to enable 'societal' types of intervention. [8] These mechanisms typically do not warrant the demonstration of individual harm or the breach of an individual's right, but can be invoked in the interest of society. Moreover, given the acknowledgment that the interest is of societal importance, the protection provided is often focused on harm prevention *ex ante* rather than mere mitigation or redress *ex post*. Accordingly, a closer look at some of the mechanisms adopted in the context of environmental law could concretise what a shift in mindset from a mere individualistic to a societal interest perspective looks like.

Just like matters relating to the internal market, the EU's environmental policy is a shared competence between the Union and member states (O'Gorman, 2013; van Calster & Reins, 2017). It addresses a diverse range of issues falling under environmental protection, from water pollution and sustainable energy to air quality. The first EU environmental intervention was primarily driven by an internal market logic: if member states maintain different environmental standards for products and services, this can constitute an obstacle to trade within the EU single market. Furthermore, the transboundary nature of the harm heightened the need for cross-border action. Interestingly for our analogy, at that time, EU primary law did not contain an explicit legal basis for environmental policy and the Union instead re-

7. It can also be pointed out that a healthy environment is one of the societal interests that can be adversely affected by the use of data-driven AI systems, given the considerable energy consumption they entail (Hao, 2019; Strubell et al., 2019).

8. Some even argue that EU environmental law focuses too much on the societal dimension of the harm, and would benefit from the recognition of an individual right to environment (O'Gorman, 2013).

lied on its general 'internal market functioning' competence (Langlet & Mahmoudi, 2016). Such a legal basis was only created at a later stage, in the margin of subsequent Treaty revisions (currently reflected in Article 11 and Articles 191-193 TFEU). EU environmental law is also heavily influenced by international law. [9]

My aim here is not to comprehensively discuss the numerous environmental directives and regulations that the EU adopted over the past decades, nor to argue that environmental law is a perfect model for (AI) regulation. Instead, I merely want to draw attention to the fact that the understanding of a problem as one that affects a societal interest also shapes its legal mechanisms. Here below I discuss three examples of protection mechanisms that merit attention.

First, rather than relying on private enforcement only [10], environmental law has also enshrined several public oversight mechanisms to ensure that the actions of public and private actors alike do not adversely impact the environment. Such public oversight has for instance taken the shape of verifying compliance with the environmental standards adopted over the years through various legislative instruments. One of the core merits of these standards concerns the creation of metrics and benchmarking of environmental performance, thereby ensuring common impact measuring methods across Europe. Environmental aspects have also been integrated into European technical standardisation more broadly (European Commission, 2004). In addition, an important oversight mechanism was introduced through the obligation to conduct an environmental impact assessment (EIA) [11]. Such assessment must be undertaken for all programmes and projects that are likely to have significant effects on the environment and is meant to ensure that the implications on the societal interest of a clean environment are taken into account before a decision is taken (Langlet & Mahmoudi, 2016). An essential element of the assessment is public participation. This participation does not hinge on the risk of individual harm, but is meant to protect the interests of society at large. Besides raising broader awareness of the issues at stake, this process also

9. For instance, and of relevance to the discussion on access to justice for societal interests, the UN-ECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters (Aarhus Convention), signed on 25 June 1998, had a substantial impact on EU environmental law.

10. As raised above, the fact that individual rights do not provide comprehensive protection against environmental harm does not diminish their importance—and the importance of private enforcement mechanisms—to tackle such harm (see e.g. Burgers, 2020). Also in the context of AI, public enforcement should be seen as complementary to (rather than substituting) private enforcement (see in this regard also Kaminski, 2019).

11. In this regard, two directives are of particular importance, namely Directive 2011/92/EU, known as the "Environmental Impact Assessment" (EIA) Directive, and Directive 2001/42/EC, known as the "Strategic Environmental Assessment" (SEA) Directive.

enhances both the transparency and accountability of decision-making processes (O'Faircheallaigh, 2010).

Second, public monitoring mechanisms have been put in place, for instance through the establishment of the European Environmental Agency (EEA). The agency's task is to provide independent information on the environment for all those developing, adopting, implementing and evaluating environmental policy—as well as to society at large. It works closely together with national environmental agencies and environmental ministries, and with the European environment information and observation network (Eionet). [12] Such public monitoring not only ensures that potential adverse effects on the environment are kept track of, but it also contributes to narrowing the knowledge gap, in light of the fact that most members of society—including public bodies—often lack the expertise to make such an analysis by themselves.

Third, a number of procedural rights were introduced with a clear 'societal' dimension, as they can be relied upon by all members of society without the need to demonstrate (a risk of) individual harm (Anton & Shelton, 2011; van Calster & Reins, 2017). Three of these—introduced through the respective three pillars of the Aarhus convention—can be pointed out in particular. First, everyone has the right to access environmental information held by public authorities, without a need to justify the access request (Krämer, 2012; Madrid, 2020). This includes information not only on the state of the environment, but also on policies or actions that were taken. In addition, public authorities are required to actively disseminate such information. Second, a right to public participation in environmental decision-making was established. Public authorities have an obligation to ensure that members of society can comment on environment-related decisions—such as the approval of projects or plans that can affect the environment—and these comments must be taken into account by the decision-maker. Information must also be provided of the final decision taken, including the reasons or justification—hence ensuring accountability and opening the door to *challengeability*. Third, an 'access to justice' right was created (Poncelet, 2012; Hadjiyianni, 2021), which provides all members of society with the right to challenge public decisions that were made without respecting the two other rights, or without complying with environmental law more generally. [13] Since this right can also be invoked by those who are indirectly im-

12. Besides the 27 EU member states, Eionet also counts Iceland, Liechtenstein, Norway, Switzerland, Turkey and the six West Balkan countries.

13. While this right has been established in theory, in practice it can however be noted that various jurisdictions - including the European Union itself - did not fully implement it, raising quite some criticism (Hadjiyianni, 2021). See also the findings and recommendations of the UN committee

pacted by the potentially harmful actions, they can be considered as *societal rights*.

5. Lessons for EU policymakers

Given the parallels identified between environmental harm and (other) harms potentially caused by the use of AI systems, policymakers aiming to address the legal protection gap arising from the overreliance on individual remedies can draw inspiration from the above mechanisms. These mechanisms exemplify how the inclusion of societal remedies can complement individual routes for redress and thereby broaden and strengthen the protection of societal interests. When translated to the context of AI, the following lessons can be drawn.

First, for those AI-applications that can adversely affect societal interests, EU policymakers should consider the introduction of public oversight and enforcement mechanisms. Rather than solely relying on those individuals who are able and willing to go to court and challenge an infringement of their individual rights, the onus can be shifted to developers and deployers of AI systems to comply with certain obligations that are subjected to public scrutiny (Yeung, 2019a). Mandatory obligations should not merely reiterate the need to comply with human rights, but should introduce specific process or outcome-oriented standards tailored to the societal interest at stake (van der Sloot & van Schendel, 2021). Moreover, they should allow organisations to demonstrate compliance with these standards, for instance through audits and certifications—without losing out of sight the limitations of algorithmic auditing (Burrell, 2016; Ananny & Crawford, 2016; Galdon Clavell et al., 2020). Public enforcement can take various forms, from a centralised to a decentralised approach, and at national or EU level—or a hybrid set-up. The sensitivity of the societal interest at stake, the expertise and resources required for effective oversight, and the legal basis that will be relied upon, are all factors that will influence the enforcement mechanism's shape. [14]

overseeing compliance with the Aarhus Convention, published in 2017, concluding that the EU did not fully comply with its obligations on access to justice, accessible at: https://unece.org/fileadmin/DAM/env/pp/compliance/CC-57/ece.mp.pp.c.1.2017.7.e.pdf.

14. It can be noted that, to some extent, inspiration can also be drawn from the GDPR (Kaminski, 2019). Despite its predominantly individualistic focus as described above, the GDPR does not entirely lay the onus on the individual, but sets out a number of mechanisms for public oversight (for instance by virtue of the role assigned to national data protection authorities) and for accountability (for instance by virtue of transparency obligations upon personal data controllers and processors) (Hoofnagle et al., 2019). While the individual, collective and societal risks raised by AI can certainly not be reduced to a personal data protection problem, the GDPR currently does function as one of the most relevant pieces of legislation to regulate AI (Hänold, 2018; Wrigley, 2018), and could provide further lessons in this context.

An additional element to consider concerns the introduction of mandatory *ex ante* impact assessments —modelled to the existing environmental impact assessments or data protection impact assessments—for AI applications that can affect societal interests. Importantly, this assessment should not only focus on the impact of the AI system's use on human rights, but also on broader societal interests such as democracy and the rule of law (CAHAI, 2020). AI developers and deployers would hence need to anticipate the potential societal harm their systems can generate as well as rationalise and justify their system's design choices in light of those risks. For transparency and accountability purposes, the impact assessments should be rendered public. Policymakers could provide further guidance as regards the assessment's format, scope and methodology. Furthermore, where AI projects can have a considerable impact on societal interests—especially in the public sector, but potentially also in private settings that have *de facto* become part of the public sphere—societal participation in the assessment process should be foreseen, analogously to the way this right exists in environmental impact assessments.

Concretely, a public administration that considers implementing an AI system in its public decision-making processes would be required to conduct a *human rights, democracy and rule of law impact assessment* prior to designing or procuring the system. This assessment would be published, for instance on the administration's website, and all interested stakeholders would be able to provide feedback on the assessment to ensure its comprehensiveness. When ultimately greenlighted, the system would need to be designed in a way that adheres to certain standards and obligations, for instance in terms of documentation and logging, verification of unjust bias, or accuracy and robustness. Finally, once deployed, the system should be regularly evaluated, independently auditable, and the necessary information to carry out such audits - or at the very least the outcomes of the independent auditor's report—should be made available to the public.

Second, EU policymakers should consider establishing a public monitoring mechanism—for instance by mandating an existing or new agency—to map and gather information on the potential adverse effects of AI systems on specific societal interests. Just like the European environmental agency is providing information about the state of the environment, there is a need to assess and disseminate independent and impartial information about the impact of AI systems on various societal interests—and in particular on society's normative and institutional infrastructure—both in the short and in the longer term. This information can not only help bridge the knowledge gap stemming from information asymmetries between AI deployers and AI affectees, but it can also inform the public on how the accu-

mulative and systematic deployment of certain AI systems might affect societal interests. In turn, societal actors—from policymakers to civil society organisations—can use this information to raise awareness of the issues at stake, improve policies, and assess whether the protective measures adopted achieve their objectives.

Importantly, monitoring mechanisms will need to rely on specific metrics and criteria to measure and assess potential (adverse) effects of certain uses of AI. Yet obtaining these metrics can be a challenge when it concerns more abstract interests such as the rule of law or societal freedom. Given that AI's societal harms often manifest themselves in an intangible manner, they herein diverge from typical environmental harms (like oil spills or air pollution) which can more easily be measured and monitored. This is also relevant when setting compliance standards, which require verifiable criteria. That said, even 'abstract' interests like the rule of law can be broken down into more concrete elements that each have an effect on its overall vitality. Inspiration can, for instance, be drawn from the World Justice Project's Rule of Law Index (World Justice Project, 2020) or the European Commission's first Rule of Law Report (European Commission, 2020b), in which the rule of law situation in various countries is assessed based on their respective systems and policies. Establishing appropriate frameworks and methodologies to map and evaluate the impact of AI systems on specific societal interests could be an important task for any future monitoring entity, and constitute a useful input for evidence-based AI policies.

Third, EU policymakers should consider strengthening procedural societal rights in the context of AI. These could include a right to access information held by public authorities on publicly used AI systems, without a need to justify the access request. As indicated above, such rights should go beyond existing access to document rights and ideally also enable citizens, researchers, civil society organisations, media and other interested parties to audit the system, all the while respecting applicable privacy laws and justifiable public interest exceptions like national security. Certain pieces of information could also be required to be made available proactively. Importantly, given the public sector's heavy reliance on private actors as regards AI systems and their underlying data, this right should also apply when AI systems are procured from private actors (Crawford & Schultz, 2019). In addition, a right to societal participation in public decision-making on AI-related projects could be established. For instance, when a public authority considers procuring, developing or deploying an AI system for a public service in a manner that risks affecting a societal interest, members of society could be given the opportu-

nity to comment on this project and to receive information—and a justification—about the decision taken. This will not only enhance accountability for the use of AI systems in public services, but can also help ensure that the potential adverse impacts on society are more comprehensively anticipated and mitigated through stakeholder participation. Last, a societal 'access to justice' right could be introduced, providing all members of society—without the need to demonstrate individual or collective harm —the right to challenge public decisions that were made without completing a comprehensive impact assessment, complying with societal participation rights, or adhering to the laws and standards applying to the use of AI systems more generally.

Finally, it should be explored whether these societal rights could also be rendered applicable when AI systems are not deployed by public authorities, but in private settings where their use nevertheless significantly affects a societal interest. [15] Certainly, the legitimate interests of private commercial actors, such as business secrets or intellectual property rights, should be respected. Simultaneously, however, an argument can be made for the introduction of societal information, participation and justice rights also in those situations where AI systems are used in a *de facto* public environment—such as social media platforms that became part of the public sphere—and hence shape society's infrastructure. [16]

It can be noted that these mechanisms need not be established through one catch-all AI regulation. AI's (societal) harms are manifold and can manifest themselves differently in specific contexts. Depending on the interest and domain at stake, mechanisms to counter them might require a tailored approach that comple-

15. Evidently, if the existence of a potential adverse impact on a societal interest dictates the applicability of the above rights, an adequate delineation of 'societal interests' becomes pressing. There is no *sui generis* definition of a 'societal interest' under European Union law. Nevertheless, primary EU legislation does offer some clues about the societal interests that the EU legal order upholds. Article 2 of the Treaty on European Union (TEU), for instance, lists the European Union's values, which both EU institutions and EU member states must comply with. These concern "respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities", which are "common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail." Of course, the fact that the EU Treaties express some of the EU's societal interests does not mean that the Union automatically has the competence to take legal action in this field. The Union's competences are limited to those specifically conferred to it by the member states and are exhaustively listed in the Treaties.

16. The concern for societal harm caused by this type of private actors seems to be one of the main drivers behind the European Commission's proposed Digital Services Act. See for instance recital 94: '*Given the importance of very large online platforms, in view of their reach and impact, their failure to comply with the specific obligations applicable to them may affect a substantial number of recipients of the services across different Member States and may cause large societal harms, while such failures may also be particularly complex to identify and address.*'

ments (existing or new) horizontal regulation. The role of existing bodies, institutions and agencies should in this regard also be considered, both as regards relevant sectors and as regards relevant interests. For instance, when it comes to monitoring AI's impact on societal interests, policymakers would need to assess whether the establishment of a new entity that bundles expertise to examine a range of interests is warranted, or whether instead reliance on (existing) specialised entities—from equality bodies and data protection authorities to sectoral agencies—is preferred.

6. The Commission's proposal for an Artificial Intelligence Act: a brief evaluation

On 21 April 2021, the European Commission published a proposal for an EU regulation on Artificial Intelligence, the so-called *Artificial Intelligence Act* (European Commission, 2021). The proposal has the dual aim of safeguarding individuals' fundamental rights against AI's adverse effects, as well as harmonising member states' rules to eliminate potential obstacles to trade on the internal market. To reach this aim, a new AI-specific legal regime is introduced, complementing rules that already apply in a technology-neutral manner (such as, for instance, the GDPR). The proposal distinguishes different categories of AI: (1) prohibited applications, which cannot be deployed unless certain exceptions apply (Title II), (2) high-risk AI systems, which need to comply with mandatory requirements prior to their placement on the market or their deployment (Title III), (3) applications that require transparency measures regardless of their risk-level (Title IV) and (4) applications that are not considered as high-risk AI systems, but may be subjected to voluntary codes of conduct that reflect similar requirements (Title IX).

Much can be said about the proposed regulation's strengths and weaknesses, yet it goes beyond the scope of this paper to conduct a detailed analysis of whether it strikes the right balance and provides adequate protection against the various harms that the use of AI systems can raise. However, bearing in mind the three types of societal protection mechanisms suggested above to counter AI's societal harm, in what follows, a brief assessment is made of the extent to which the proposal takes these mechanisms into consideration.

First, the proposal can be commended for introducing a public oversight and enforcement framework for high-risk AI applications. [17] This signals that the need for

17. It can be noted that this is in line with the proposal for a Digital Services Act, which also covers the use of certain AI systems on online platforms and which likewise establishes a public oversight mechanism, thus recognising the societal interests at stake. Interestingly, in the recitals of this pro-

adequate safeguards against AI's risks is a matter of importance to society at large, and not only for the potentially harmed individuals or collectives. The proposal thus bridges an important protection gap by shifting the burden from individuals to independent national supervisory authorities to ensure that a set of essential rules are respected. In this regard, the introduction of prohibitions and *ex ante* requirements in terms of data quality, transparency and accountability for high-risk AI systems, the possibility to conduct independent audits, and the threat of substantive fines are important contributors. Moreover, a publicly accessible EU database of stand-alone high-risk AI systems will be established, managed by the European Commission, which can increase public transparency. At the same time, however, some caveats can be made.

As of today, many of the proposed requirements—for instance in terms of data quality or human oversight—still lack uniform implementation standards, which renders it difficult to assess their compliance in a non-arbitrary fashion. This is especially important considering that enforcement is organised at the national level, which raises the risk that different protection standards may be applied by the different national supervisory authorities, and that citizens across the EU may not be equally protected [18] (in addition to the risk that not all authorities will be equipped with sufficient resources to fulfil their task, reminiscent of the problems faced in the context of the GDPR). [19] It can also be questioned whether the list of high-risk AI applications is sufficiently comprehensive, and whether some of those applications should not rather undergo an *ex ante* authorisation process by an independent authority, rather than solely facing the possibility of *ex post* control—especially in public contexts. In addition, while conducting a risk assessment seems to be required for high-risk AI systems, these assessments are not rendered public, nor is there a possibility for society to provide feedback thereon. Only national authorities seem to be able to request access to the AI systems' (assessment) documentation. Finally, neither citizens nor civil society organisations are granted the right to file a complaint with the national supervisory authority in case they suspect non-compliance with the regulation. In other words, the shift from

posal, explicit use is made of terms like 'societal harm', 'societal risks' and 'societal concerns', albeit without precisely defining what is meant therewith.

18. It can however be noted that a European Artificial Intelligence Board will be established with representatives of the national supervisory authorities, which will serve *inter alia* to "contribute to uniform administrative practices in the member states", as per the proposed Article 58 of the regulation.

19. Consider in this regard the report published by Access Now in May 2020, in which the lack of adequate resources for certain data protection authorities is highlighted. The report is accessible at: https://www.accessnow.org/cms/assets/uploads/2020/05/Two-Years-Under-GDPR.pdf.

private to public enforcement might arguably have been taken somewhat too far since the regulation appears to rely entirely on national authorities without envisaging any role for society. There is, however, no reason why both could not act in a complementary or cooperative manner.

Second, it can be noted that the proposed regulation does not explicitly provide for a public monitoring mechanism to map and disseminate independent information on the potential (adverse) effects of AI systems on specific societal interests, such as equality or the rule of law. The establishment of *market surveillance authorities* is proposed to monitor compliance with the regulation's requirements, yet at first sight these authorities are not meant to research, or collect information on, the societal impact that the implementation of AI—for instance in public infrastructures—could generate over the longer term. And while the aforementioned EU database could certainly increase transparency about existing stand-alone high-risk AI systems, it still does not fit the bill of the second societal protection mechanism outlined in the chapter above. At this stage, it remains to be seen whether the proposed *European Artificial Intelligence Board* – which will be chaired by the Commission and composed of the national supervisory authorities as well as the European Data Protection Supervisor – could play a role in this regard. [20]

Third, as regards the introduction of procedural rights with a societal dimension, the proposed regulation is entirely silent. In fact, the drafters of the proposal seem to have been very careful *not* to establish any new rights in the regulation, and only impose obligations. This means that, if an individual wishes to challenge the deployment of an AI system that breaches the requirements of the regulation or adversely affects a societal interest, she will still need to prove individual harm. As already mentioned above, she will also not be able to lodge a complaint with the national supervisory authority to investigate the potentially problematic practice—unless the national authority decides to provide this possibility on its own motion. Besides the lack of an access to justice right, the proposed regulation also does not provide for an access to the information right or a right to societal participation in public decision-making on AI related projects.

Finally, and more generally, it is clear that the proposal remains imbued by concerns relating almost exclusively to individual harm, and seems to overlook the need for protection against AI's societal harms. This manifests itself in at least four ways: (1) the prohibition of certain manipulative AI practices in Title II are careful-

20. While not explicitly mentioned in the proposal, it can be noted that the European Commission already announced its intention to establish an expert group that could contribute to the European Artificial Intelligence Board's work.

ly delineated so as to require individual harm—and in particular, "*physical or psychological harm*" [21]—thus excluding the prohibition of AI systems that cause societal harm; (2) as regards the high-risk systems of Title III, their risk is only assessed by looking at the system in an isolated manner, rather than taking into account the cumulative or long term effects that its widespread, repetitive or systemic use can have; (3) the Commission can only add high-risk AI systems to the list of Annex III in case these systems pose "*a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights*" [22], thus excluding risks to societal interests; (4) certain AI practices that are highly problematic from a societal point of view—such as AI-enabled emotion recognition [23] or the AI-enabled biometric categorisation of persons [24]—are currently not even considered as high-risk.

In sum, while the Commission's proposal constitutes a firm step towards enhanced protection against AI's adverse effects, it does not fully reflect the societal dimension thereof. Moreover, it does not provide a role for societal actors to challenge the way in which AI systems can cause societal harm, whether by ensuring the collection of and access to relevant information, granting the possibility to give feedback and participate in public decision-making, or facilitating access to justice. One can wonder whether this is not a missed opportunity, given the importance of the interests at stake. At the same time, as mentioned previously, mechanisms to counter AI's societal interests need not be tackled in a single regulation, but may well be established or strengthened through various actions. Moreover, it should be kept in mind that this proposal is subject to change, as both the European Parliament and the Council of the European Union will now have the opportunity to propose alterations. [25] In other words, there is still scope for further improvement, and it can be hoped that EU policymakers will place the societal dimension of AI's harm more prominently on their radar.

21. See in this regard Article 5(1)(a) and (b) of the proposal.
22. See in this regard Article 7 of the proposed regulation.
23. The only exception concerns AI systems intended to be used by public authorities "*as polygraphs and similar tools or to detect the emotional state of a natural person*" either in the context of migration, asylum and border control management, or in the context of law enforcement. See in this regard Annex III, point 6(b) and 7(a).
24. The insertion of "*biometric categorisation of natural persons*" in the *title* of Annex III, point 1, does appear to indicate that the Commission could include such systems as 'high-risk' upon a revision of this Annex at a later stage, if this inclusion can be justified pursuant to Article 7 of the proposal.
25. The proposed regulation, relying on articles 16 and 114 of the Treaty of the Functioning of the European Union as its legal basis, will in principle need to be adopted through the ordinary legislative procedure.

7. Conclusions

The development and use of AI applications risks not only causing individual and collective harm, but also societal harm—or wrongful setbacks to societal interests. The societal impact of AI systems is increasingly acknowledged, yet the fact that it cannot always be equated with the sum of individual harms often still remains overlooked. As a consequence, policymakers aiming to analyse legal gaps in the legislative framework applicable to AI systems and other data-driven technologies often fail to account for this discrepancy, and thereby also pay insufficient attention to the legal protection gap arising from an overreliance on individual remedies to safeguard adversely affected societal interests.

I argued above that, even in those circumstances when a specific human right could be invoked in a context where societal harm ensues—for instance because an individual's right to data protection or right to non-discrimination has been infringed—there is still a risk that this opportunity runs aground. First, the individual may not know that a right was infringed, as many AI systems operate in a non-transparent manner (*the knowledge gap problem*). Second, the individual could be aware of and opposed to the practice, but the individual harm may be disproportionately small compared to the effort and resources it would take to challenge the practice in court—even if the societal harm arising therefrom may be significant (*the threshold problem*). Third, the individual may have consented to the use of her personal data or to the use of the AI system, for instance in exchange for a particular service, and might be unaware of—or unbothered by—the fact that this action may subsequently cause societal harm (*the egocentrism problem*).

To bridge the legal protection gap that leaves societal interests vulnerable to AI-enabled harm, I argue that policymakers should shift their perspective from an individualistic one to a societal one, by recognising the societal interests threatened by the use of AI as *sui generis* interests. This means that AI's adverse effects on societal interests should not only be included in AI-related harm analyses, but also in proposals for new legislation. Such legislation could help counter societal harm more effectively by introducing legal remedies with a societal dimension. Drawing on the domain of environmental law, which aims to protect one of the most recognised societal interests, several such mechanisms were identified. While the European Commission's proposal for an AI regulation seems to head into the right direction by introducing a public oversight structure for certain AI applications, the proposal does not incorporate the proposed *societal* mechanisms to bridge the legal protection gap as regards AI's societal harm and could benefit from further improvement.

To conclude this article, five caveats merit being spelled out at this stage, and can at the same time be read as a future research agenda. First, it has been stressed that a very diverse set of societal interests can be adversely impacted by the use of AI systems. Hence, a one-size-fits-all approach to tackle AI's societal harms is unlikely to succeed. While some mechanisms can contribute to the protection of several societal interests, other interests may require more tailored solutions.

Second, it should be borne in mind that AI systems are but one technology amongst many others, and that the societal harm their use might generate will not always be unique to their features. Indeed, many societal risks are not specific to AI, and it may be counterproductive to treat them as such (Smuha, 2021). As a consequence, in some instances, introducing societal remedies that are not necessarily AI-specific—while nevertheless covering the harm they raise—could be more effective.

Third, the above overview of environmental legislation only focused on three types of legal mechanisms and should not be understood as an exhaustive parallel; other mechanisms could also provide inspiration. At the same time, the broad nature of the above overview also means it did not contain an analysis of how these mechanisms fall short, yet it is certainly not argued that they are infallible, nor is it argued that the analogy between AI's societal harm and environmental harm always stands. [26] Instead, my primary aim is drawing attention to their underlying *societal* logic.

Fourth, the specific shape and scope of potentially new legal mechanisms, as well as the feasibility of introducing them at the EU level, will rely on the legal basis that can be invoked—which will in turn also determine if they can withstand the subsidiarity and proportionality test. As noted above, at this stage, it remains to be seen whether the legal basis that the Commission's regulatory proposal for AI relies on, is in fact appropriate to counter AI's societal impact beyond interests related to personal data protection issues or internal market issues.

Last, the role of EU member states—and their potential desires to maintain jurisdictional competence to assess and address AI's impact on societal interests—will also need to be considered. Although, in principle, all member states adhere to the values set out in Article 2 TEU, in practice, their cultural and political interpretations of these values can at times diverge. The degree of this divergence will be an

26. Consider, for instance, the difference in terms of tangibility between societal harms raised by AI and societal harms of environmental nature—and hence also in terms of measurability, quantifiability and verifiability.

important factor in the process to adopt EU-wide policies in this domain.

Regardless of these caveats, societal interests will not be adequately protected as long as policymakers do not look beyond the individual when analysing the shortcomings of the current legal framework. They should hence shift their perspective by including an assessment of AI's potential to cause societal harm, and ensure effective governance mechanisms to tackle it. This paper aims to contribute to such assessment.

References

Ala-Pietilä, P., & Smuha, N. A. (2021). A Framework for Global Cooperation on Artificial Intelligence and Its Governance. In B. Braunschweig & M. Ghallab (Eds.), *Reflections on Artificial Intelligence for Humanity* (pp. 237–265). Springer International Publishing. https://doi.org/10.1007/978-3-030-69128-8_15

AlgorithmWatch. (2020). *Automating Society Report* [Report]. AlgorithmWatch; Bertelsmann Stiftung. https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/10/Automating-Society-Report-2020.pdf

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Anton, D. K., & Shelton, D. (2011). *Environmental Protection and Human Rights*. Cambridge University Press. https://doi.org/10.1017/CBO9780511974571

Bamberger, K. A., & Mayse, A. (2020). Privacy in Society: Jewish Law Insights for the Age of Big Data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3731770

Barrett, L. (2018). Model(Ing) Privacy: Empirical Approaches to Privacy Law & Governance. *Santa Clara High Tech Law Journal*, *35*(1). https://digitalcommons.law.scu.edu/chtlj/vol35/iss1/1

Bayamlioglu, E. (2018). Contesting Automated Decisions: *European Data Protection Law Review*, *4*(4), 433–446. https://doi.org/10.21552/edpl/2018/4/6

Bayamlıoğlu, E., & Leenes, R. (2018). The 'rule of law' implications of data-driven decision-making: A techno-regulatory perspective. *Law, Innovation and Technology*, *10*(2), 295–313. https://doi.org/10.1080/17579961.2018.1527475

Bietti, E. (2020). Consent as a Free Pass: Platform Power and the Limits of the Informational Turn. *Pace Law Review*, *40*(1), 310–398. https://digitalcommons.pace.edu/plr/vol40/iss1/7

Binns, R. (2018). Algorithmic Accountability and Public Reason. *Philosophy & Technology*, *31*(4), 543–556. https://doi.org/10.1007/s13347-017-0263-5

Brkan, M. (2019). Artificial Intelligence and Democracy: *Delphi - Interdisciplinary Review of Emerging Technologies*, *2*(2), 66–71. https://doi.org/10.21552/delphi/2019/2/4

Brownsword, R. (2016). Technological management and the Rule of Law. *Law, Innovation and*

Technology, *8*(1), 100–140. https://doi.org/10.1080/17579961.2016.1161891

Buchholtz, G. (2020). Artificial Intelligence and Legal Tech: Challenges to the Rule of Law. In T. Wischmeyer & T. Rademacher (Eds.), *Regulating Artificial Intelligence* (pp. 175–198). Springer International Publishing. https://doi.org/10.1007/978-3-030-32361-5_8

Burgers, L. E. (2020). *Justitia, the People's Power and Mother Earth: Democratic legitimacy of judicial law-making in European private law cases on climate change* [PhD Thesis, University of Amsterdam]. https://pure.uva.nl/ws/files/52346648/Front_matter.pdf

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 205395171562251. https://doi.org/10.1177/2053951715622512

CAHAI (Council of Europe Ad Hoc Committee on Artificial Intelligence). (2020). *Feasibility Study* [CAHAI(2020)23]. Council of Europe. https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da

Cate, F. H., & Mayer-Schonberger, V. (2013). Notice and consent in a world of Big Data. *International Data Privacy Law*, *3*(2), 67–73. https://doi.org/10.1093/idpl/ipt005

Cohen, J. E. (2017). Affording Fundamental Rights: A Provocation Inspired by Mireille Hildebrandt. *Critical Analysis of Law*, *4*(1), 76–90. https://scholarship.law.georgetown.edu/facpub/1964

Cohen, J. E. (2019). *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford University Press. https://doi.org/10.1093/oso/9780190246693.001.0001

Conaghan, J. (2002). Law, harm and redress: A feminist perspective. *Legal Studies*, *22*(3), 319–339. https://doi.org/10.1111/j.1748-121X.2002.tb00196.x

Council Europe. (2019). *Terms of reference for the Ad hoc Committee on Artificial Intelligence (CAHAI)*.

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Sánchez, A. N., Raji, D., Rankin, J. L., Richardson, R., Schultz, J., West, S. M., & Whittaker, M. (2019). *AI Now 2019 Report* [Report]. AI Now Institute. https://ainowinstitute.org/AI_Now_2019_Report.pdf

Crawford, K., & Schultz, J. (2019). AI systems as state actors. *Columbia Law Review*, *119*(7), 1941–1972. https://columbialawreview.org/content/ai-systems-as-state-actors/

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer International Publishing. https://doi.org/10.1007/978-3-030-30371-6

Durkheim, E. (1925). *L'éducation morale*. Alcan.

European Commission. (2004). *Communication from the Commission to the Council, the European Parliament and the European Economic and Social Committee—Integration of Environmental Aspects into European Standardisation {SEC(2004)206}*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52004DC0130

European Commission. (2020a). *Inception Impact Assessment on the Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence*.

European Commission. (2020b). *Rule of Law Report—The Rule of Law Situation in the European Union*.

European Commission. (2020c). *White Paper on Artificial Intelligence—A European approach to excellence and trust* (White Paper COM(2020) 65 final). European Commission. https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52020DC0065

European Commission. (2021). *Proposal for a Regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*.

European Union Agency for Fundamental Rights. (2020). *Getting the future right – Artificial intelligence and fundamental rights* [Report]. Publications Office of the European Union. https://doi.org/10.2811/774118

Feinberg, J. (1984). Harm to Others. In *The Moral Limits of the Criminal Law—Volume 1: Harm to Others*. Oxford University Press.

Galdon Clavell, G., Martín Zamorano, M., Castillo, C., Smith, O., & Matic, A. (2020). Auditing Algorithms: On Lessons Learned and the Risks of Data Minimization. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 265–271. https://doi.org/10.1145/3375627.3375852

Gebru, T. (2020). Race and Gender. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (pp. 251–269). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.16

Hadjiyianni, I. (2021). Judicial protection and the environment in the EU legal order: Missing pieces for a complete puzzle of legal remedies. *Common Market Law Review*, *58*(3), 777–812. https://kluwerlawonline.com/journalarticle/Common+Market+Law+Review/58.3/COLA2021050

Hänold, S. (2018). Profiling and Automated Decision-Making: Legal Implications and Shortcomings. In M. Corrales, M. Fenwick, & N. Forgó (Eds.), *Robotics, AI and the Future of Law* (pp. 123–153). Springer Singapore. https://doi.org/10.1007/978-981-13-2874-9_6

Hao, K. (2019, June 6). Training a single AI model can emit as much carbon as five cars in their lifetimes. *MIT Technology Review*. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Hao, K. (2021, March 11). He Got Facebook Hooked on AI. Now He Can't Fix Its Misinformation Addiction. *MIT Technology Review*. https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/

Hasselbalch, G. (2019). Making sense of data ethics. The powers behind the data ethics debate in European policymaking. *Internet Policy Review*, *8*(2). https://doi.org/10.14763/2019.2.1401

High-Level Expert Group AI. (2019). *A definition of AI: Main capabilities and scientific disciplines*. European Commission. https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI* [Report]. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Hildebrandt, M. (2015). *Smart Technologies and the End(s) of Law*. Edward Elgar. https://doi.org/10.4337/9781849808774

Hildebrandt, M. (2018). Algorithmic regulation and the rule of law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128), 20170355. https://doi.org/10.1098/rsta.2017.0355

Hoofnagle, C. J., van der Sloot, B., & Borgesius, F. Z. (2019). The European Union general data protection regulation: What it is and what it means. *Information & Communications Technology Law*, *28*(1), 65–98. https://doi.org/10.1080/13600834.2019.1573501

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, *51*(8), 56–59. https://doi.org/10.1109/MC.2018.3191268

Jørgensen, R. F. (Ed.). (2019). *Human Rights in the Age of Platforms*. The MIT Press. https://doi.org/10.7551/mitpress/11304.001.0001

Kaminski, M. E. (2019). Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability. *Southern California Law Review*, *92*(6), 1529–1616. https://southerncalifornialawreview.com/2019/09/01/binary-governance-lessons-from-the-gdrps-approach-to-algorithmic-accountability-article-by-margot-e-kaminski/

Kernohan, A. (1993). Accumulative Harms and the Interpretation of the Harm Principle. *Social Theory and Practice*, *19*(1), 51–72. https://doi.org/10.5840/soctheorpract19931912

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78–85. https://doi.org/10.1145/2500873

Krämer, L. (2012). Transnational Access to Environmental Information. *Transnational Environmental Law*, *1*(1), 95–104. https://doi.org/10.1017/S2047102511000070

Kutz, C. (2000). *Complicity: Ethics and Law for a Collective Age* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511663758

Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (Eds.). (2014). *Privacy, Big Data, and the Public Good—Frameworks for Engagement*. Cambridge University Press. https://doi.org/10.1017/CBO9781107590205

Langlet, D., & Mahmoudi, S. (2016). EU Environmental Law and Policy. In *EU Environmental Law and Policy*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198753926.001.0001

Liu, H.-W., Lin, C.-F., & Chen, Y.-J. (2019). Beyond State v Loomis: Artificial intelligence, government algorithmization and accountability. *International Journal of Law and Information Technology*, *27*(2), 122–141. https://doi.org/10.1093/ijlit/eaz001

Madrid, J. Z. (2020). *Access to Environmental Information held by the Private Sector under International, European and Comparative Law*. KU Leuven.

Mittelstadt, B. (2017). From Individual to Group Privacy in Big Data Analytics. *Philosophy & Technology*, *30*(4), 475–494. https://doi.org/10.1007/s13347-017-0253-7

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 205395171667967. https://doi.org/10.1177/2053951716679679

Muller, C. (2020). The Impact of AI on Human Rights, Democracy and the Rule of Law. In CAHAI Secretariat (Ed.), *Towards regulation of AI systems ((CAHAI(2020)06)* (pp. 23–31). Council of Europe. https://www.coe.int/en/web/artificial-intelligence/cahai

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics*, *2*(1), 25–42. https://doi.org/10.1007/BF02639315

O'Faircheallaigh, C. (2010). Public participation and environmental impact assessment: Purposes, implications, and lessons for public policy making. *Environmental Impact Assessment Review*, *30*(1), 19–27. https://doi.org/10.1016/j.eiar.2009.05.001

O'Gorman, R. (2013). The Case for Enshrining a Right to Environment within EU Law. *European*

Public Law, *19*(3), 583–604.

O'Neil, C. (2017). *Weapons of Math Destruction*. Penguin Books Ltd.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Poncelet, C. (2012). Access to Justice in Environmental Matters—Does the European Union Comply with its Obligations? *Journal of Environmental Law*, *24*(2), 287–309. https://doi.org/10.1093/jel/eqs004

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.

Schermer, B., Custers, B., & van der Hof, S. (2014). The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection. *Ethics and Information Technology*, *16*(2), 171–182. https://doi.org/10.1007/s10676-014-9343-8

Simon, T. W. (1995). *Democracy and Social Injustice: Law, Politics, and Philosophy*. Rowman & Littlefield.

Smuha, N. A. (2020). Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea. *Philosophy & Technology*. https://doi.org/10.1007/s13347-020-00403-w

Smuha, N. A. (2021). From a 'race to AI' to a 'race to AI regulation': Regulatory competition for artificial intelligence. *Law, Innovation and Technology*, *13*(1), 57–84. https://doi.org/10.1080/17579961.2021.1898300

Solow-Niederman, A. (2019). Administering Artificial Intelligence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3495725

Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP* [Preprint]. ArXiv. http://arxiv.org/abs/1906.02243

Tamo-Larrieux, A., Mayer, S., & Zihlmann, Z. (2020). *Not Hardcoding but Softcoding Privacy*. https://www.alexandria.unisg.ch/262254/

Taylor, L., Floridi, L., & van der Sloot, B. (Eds.). (2017). *Group Privacy: New Challenges of Data Technologies*. Springer International Publishing. https://doi.org/10.1007/978-3-319-46608-8

Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, *2*(1), 10–12. https://doi.org/10.1038/s42256-019-0136-y

Thompson, D. F. (1980). Moral Responsibility of Public Officials: The Problem of Many Hands. *American Political Science Review*, *74*(4), 905–916. https://doi.org/10.2307/1954312

UNESCO. (2020). *Outcome document: First draft of the Recommendation on the Ethics of Artificial Intelligence (SHS/BIO/AHEG-AI/2020/4 REV.2). Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000373434

van Alsenoy, B., Kosta, E., & Dumortier, J. (2014). Privacy notices versus informational self-determination: Minding the gap. *International Review of Law, Computers & Technology*, *28*(2), 185–203. https://doi.org/10.1080/13600869.2013.812594

van Calster, G., & Reins, L. (2017). *EU Environmental Law*. Edward Elgar Publishing.

van de Poel, I., Nihlén Fahlquist, J., Doorn, N., Zwart, S., & Royakkers, L. (2012). The Problem of

Many Hands: Climate Change as an Example. *Science and Engineering Ethics*, *18*(1), 49–67. https://doi.org/10.1007/s11948-011-9276-0

van der Sloot, B. (2017). *Privacy as Virtue: Moving Beyond the Individual in the Age of Big Data* (1st ed.). Intersentia. https://doi.org/10.1017/9781780686592

van der Sloot, B., & van Schendel, S. (2021). Procedural law for the data-driven society. *Information & Communications Technology Law*, *30*(3), 304–332. https://doi.org/10.1080/13600834.2021.1876331

Véliz, C. (2020). *Privacy is Power*. Bantam Press.

Viljoen, S. (2020). Democratic Data: A Relational Theory For Data Governance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3727562

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121 136. https://www.jstor.org/stable/20024652

World Justice Project. (2020). *Rule of Law Index 2020* [Report]. World Justice Project. https://worldjusticeproject.org/sites/default/files/documents/WJP-ROLI-2020-Online_0.pdf

Wrigley, S. (2018). Taming Artificial Intelligence: "Bots," the GDPR and Regulatory Approaches. In M. Corrales, M. Fenwick, & N. Forgó (Eds.), *Robotics, AI and the Future of Law* (pp. 183–208). Springer. https://doi.org/10.1007/978-981-13-2874-9_8

Yeung, K. (2019a). *Responsibility and AI - A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework* (Study DGI)2019)05). Council of Europe. https://rm.coe.int/responsability-and-ai-en/168097d9c5

Yeung, K. (2019b). Why Worry about Decision-Making by Machine? In K. Yeung, *Algorithmic Regulation* (pp. 21–48). Oxford University Press. https://doi.org/10.1093/oso/9780198838494.003.0002

Yeung, K., Howes, A., & Pogrebna, G. (2020). AI Governance by Human Rights–Centered Design, Deliberation, and Oversight: An End to Ethics Washing. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (pp. 75–106). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.5

Zalnieriute, M., Moses, L. B., & Williams, G. (2019). The Rule of Law and Automation of Government Decision-Making. *The Modern Law Review*, *82*(3), 425–455. https://doi.org/10.1111/1468-2230.12412

Zuiderveen Borgesius, F. J. (2015). Behavioural Sciences and the Regulation of Privacy on the Internet. In A. Alemanno & A.-L. Sibony (Eds.), *Nudge and the Law: A European Perspective* (pp. 179–208). Hart Publishing. https://doi.org/10.5040/9781474203463

Zuiderveen Borgesius, F. J. (2018). *Discrimination, artificial intelligence, and algorithmic decision-making* [Study]. Council of Europe, Directorate General of Democracy. https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B., & De Vreese, C. (2018). Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review*, *14*(1), 82. https://doi.org/10.18352/ulr.420

Published by


in cooperation with