Accountable AI: It Takes Two to Tango

Jorge Constantino

Abstract

This Chapter argues that accountable artificial intelligence (AI) requires examining the role of humans in AI development and deployment. Hence, it discusses the importance of addressing the obligations of deployers and developers of AI systems to achieve accountable AI. The EU AI Act has implemented measures such as transparency or technical obligations to achieve such accountability. Similarly, it has implemented human oversight requirements outlined in Arts. 14 and 26 against high-risk AI systems. Some scholars and practitioners may argue that Art. 14 only applies to developers of AI systems. However, we understand that human oversight requirements govern both actors. Human oversight cannot be applied in isolation by requiring compliance of only one party. Otherwise, it would defeat the purpose of adding human control features to prevent AI systems from harming fundamental rights. Based on this perspective, we propose that (at least) two actors are required to make accountable AI more tangible. Nonetheless, we are conscious that this legislation is in its infancy, and only time will tell how human oversight obligations (Arts. 14 and 26) are to be applied – whether in isolation or in conjunction.

1. An introduction to AI systems

Artificial intelligence (AI) is currently used in the public and private sectors in such fields as policing, the judicial system, employment, taxes and finances, retailers, media, and entertainment (Maclure, 2020, pp. 2–3; Sipola et al, 2024, p. 5). The definition or conceptualisation of AI is far from settled (Kuziemski and Misuraca, 2020, p. 2). For instance, AI may be simply defined as computers or machines showing human-like intelligence (Simmons and Chappell, 1988, p. 14) (DK, 2023, p. 7). Alternatively, some academics have described AI as the umbrella term that refers to a set of algorithmic models, methods, or instructions given to a computer system to simulate human intelligence (Köchling and Wehner, 2020, p. 798; Muthukrishnan et al, 2020, p. 393). Thus, for the purpose of this study, it may be helpful to refer to the definition of AI found in the EU's AI Act (Regulation 2024/1689), which refers to a machine-learning system designed with different levels of autonomy that requires inputs to produce outputs influencing the physical or virtual environment with which they interact.[1] Similarly, according to Muthukrishnan et al (2020, pp. 394–395), machine learning is a subfield of AI that involves some form of learning using data samples.

Following the AI Act's proposed definition, we may agree that AI comes in different *forms* and *shapes*; for example, machine learning, not being fully autonomous, requires human intervention to learn from algorithms or datasets and be able to solve tasks (Kowalski, 1979, p. 424; Hill, 2016, pp. 35–36, 58). Thus, while some AI advocates may preach that AI resembles (or even surpasses) human intelligence, the reality is that AI (or, at least, machine learning) is not always fully autonomous. We may argue that human intervention will always be needed for an AI system to come alive and work as an "intelligent" thing (Lennox, 2020, pp. 53–61). Nonetheless, the "intelligence" of such systems is not the focus of this Chapter. Rather, our argument is that to examine accountable AI systems, it is necessary to analyse the human factor in the process of their development and deployment. For instance, what would be the cause and result of AI failures: designers, deployers, or the machine itself? (Edwards, Schafer and Harbinja, 2020, p. 310) Thus, to guide our analysis, we have formulated the following question: "what accountability measures has the European AI Act implemented to protect fundamental rights against harmful AI?" In the following paragraphs, we attempt to provide some answers, arguing that, at this very stage, machines or AI systems have no legal capacity to be held accountable themselves. Thus, at least for now, accountable AI requires examining the roles of two human actors: developers and deployers (Constantino, 2022, p. 2). Thus, we argue that it takes two to tango in accountable AI.

2. AI systems in our societies: good and bad AI?

AI systems can positively impact our societies (Henao, 2021), help fight crime (Eligon and Williams, 2015), assist in having more efficient services

[1] This is an adapted definition from the European AI Act. Please refer to Art. 3 EU AI Act for a full definition.

(Linden, 2021, p. 2), be more cost-effective (Le Sueur, 2015, pp. 3, 18), and even – as some have argued – offer less discriminatory results compared to human decision-makers (Chander, 2017, p. 1027; Clifford, 2017, p. 94; Hacker, 2018, p. 3). Similarly, AI systems can be used to establish risk scores regarding tax and welfare fraud and unlawful immigration (Maclure, 2020, pp. 2–3).

AI systems are currently used in the public sector to make the bureaucratic system more responsive and simpler to citizens seeking social security assistance or lodging tax returns (Le Sueur, 2015, p. 3). From the broad use (or deployment) of AI systems in public sectors in different countries, we may come across two contested cases of their deployment in government: the Dutch experience with the *System Risico Indicatie* (SyRI) and the Australian experience with *Robodebt*. In the former, the SyRI deployed AI tools to identify citizens who may have potentially committed or may represent a risk of committing social security fraud (Wisman, 2020). SyRI had the legal and technological power to link and analyse citizens' personal data concerning work data, administrative fines, tax data, real estate and personal assets, housing, civic integration data, education data, social benefits, and subsidies (NJCM et al v. The Dutch State, 2020, p. 4.17). SyRI had the task of collecting and analysing citizens' data, preparing reports based on profiling people and providing a risk score regarding certain citizens, thereby warning the Dutch authorities of potential social services fraud (NJCM et al. v. The Dutch State, 2020, p. 4.17). As defended by the Dutch government, the implementation of the SyRI provided an advantage in targeting those who were committing fraud, and thereby damaging the country's economy and social security service (NJCM et al v. The Dutch State, 2020, p. 6.3, 6.76). However, the SyRI was found to be unlawful for numerous reasons, such as breaching human rights and privacy laws and the lack of transparency on the part of the Dutch government to reveal the inner workings and purpose of the AI system in use (NJCM et al v. The Dutch State, 2020, p. 6.5, 6.27, 6.32, 6.41).

A similar case occurred in Australia in 2016; the federal government rolled out an AI system labelled *Robodebt* to detect citizens who apparently received social security overpayments (Whiteford, 2021, p. 340). The *Robodebt* system collected data from former and current welfare beneficiaries and compared it against their annual tax income assessment to automatically ascertain any overpayment (Whiteford, 2021, pp. 341–342). Unfortunately, the automated system was built with inaccurate algorithms, leading to miscalculations. *Robodebt* shifted the burden of proof onto citizens to demonstrate they were not overpaid; if a citizen could not prove that the automated system was incorrect, the system would generate a debt against that citizen (Human Rights Law Centre, 2021). In November 2019, the Australian Federal Court ruled that the *Robodebt* system was unlawful and ordered the Australian government to return the money unlawfully collected to recipients of welfare payments (Whiteford, 2021, p. 347). The Court held that the Australian government failed in its duty to citizens to oversee the correct functioning of *Robodebt*, and that the government had blindly relied upon the automated system without putting in place any human intervention to verify the accuracy of the AI system (Human Rights Law Centre, 2021).

AI systems are also being deployed in the private sector across different markets. For example, financial organisations use AI systems to assign risk score credit to applicants before deciding on whether to grant loans (Pasquale, 2015, p. 1; Chander, 2017, p. 1024). Amazon built an AI system to assist its human resources department in choosing the top five candidates out of hundreds of applicants (Winick, 2018). However, it has been reported that Amazon realised that its AI system negatively discriminated against women and preferred men as suitable candidates (Winick, 2018). Google offers AI systems that can help users collect, categorise, and automatically tag uploaded photos to simplify users' lives (Dougherty, 2015). However, it has been reported that Google’s face recognition algorithm mistakenly labelled black people as gorillas due to insufficient training data on recognising black faces (Hacker, 2018, p. 7). Furthermore, in recent years, researchers have developed AI-supported care robots to monitor the elderly and assist with such basic tasks as reducing loneliness or ensuring that prescriptions are taken at the right time (Johansson-Pajala and Gustafsson, 2020, p. 167). For example, the robot PARO assists the elderly with dementia and Alzheimer’s (Kelly et al, 2021). It is claimed that PARO can help in reducing stress and anxiety (Kang et al, 2020) and can detect patients’ body temperature (Kang et al, 2020). However, as these medical devices are part of the Internet of Things (IoT), their functionality depends on data exchanges to connect with other compatible networks to support their operation (Ray, 2016, pp. 9489–9491). Thus, these medical devices are unfortunately exposed to cybersecurity vulnerabilities, such as patients’ data being stolen by cybercriminals (Drukarch, Calleja and Fosch Villaronga, 2023, pp. 15–16).

Further to the above, there are numerous other examples of AI developments and deployments covering various applications across different sectors, such as AI systems for intelligence, military, and national security purposes (Constantino and van der Linden, 2024, pp. 1–5; Barzashka, 2023, pp. 26–27), video surveillance through smart technology in the workplace to monitor production, safety, and control of employees entering and leaving the workplace (Rosenblat, Kneese and Boyd, 2014, pp. 2–3, 7–10). However, the above examples may be enough to illustrate the complexities and risks of AI systems in our societies, whether in Europe, the US, or Australia.

We can observe that AI systems may help fight fraud or crime. However, if an AI system is developed with inaccurate data or inherent bias from human developers, it is likely to pose a risk of discrimination or unfairness during its deployment (Edwards, Schafer and Harbinja, 2020, p. 238). For example, inaccurate data that feeds AI systems can contain prejudicial stigmas against certain groups of people, can contain racial discrimination, and can occasionally be tainted by unlawful practices (Richardson Schultz, and Crawford, 2019, p. 15). Historical data provided to AI systems can lead to discriminatory results, such as insufficient data or lack of robust data (Edwards, Schafer and Harbinja, 2020, p. 238; Chander, 2017, p. 1036). AI systems can capture and reproduce negative discrimination in their outputs and be contaminated by training data and natural operations in the real world, thereby leading to the reproduction of real-world negative discrimination towards citizens (Hacker, 2018, pp. 34–35). Moreover, even when AI systems are designed in a “neutral” manner, there is no guarantee that they will behave flawlessly (Hacker, 2018, p. 11). This begs the question, how lawful are these AI systems? Are faulty AI systems the result of reckless programming or poor deployment? (Richardson Schultz and Crawford, 2019, pp. 15, 48).

From the examples provided, we may argue that the SyRI reinforced further disparity and discrimination against those living in poverty and needing welfare assistance (Appelman, Ó Fathaigh and van Hoboken, 2021, p. 341). Faulty AI systems can harm society, and particularly its most vulnerable members (Maclure, 2020, p. 1044). Similarly, an AI system without the proper supervision of capable and willing humans is also likely to pose a risk to citizens who come into contact with it. For example, appropriate human oversight measures in the *Robodebt* system may have prevented fatal consequences that had endangered human life (Whiteford, 2021, p. 341). Without adequate measures to develop and deploy AI systems that support the core of human dignity, we may be left in a society where AI systems are employed to oppress and target vulnerable citizens (Whiteford, 2021, p. 356).

Furthermore, AI systems deployed in our societies may pose other risks to fundamental rights, such as the right to privacy and data protection. For instance, surveillance in the workplace may be used for ill, such as in the harassment or exploitation of employees (Sykes, 2000). Deploying invasive technologies affects employees' right to privacy, even if deployed inside the workplace, because the employee is not expected to be monitored in the workplace (European Data Protection Board (EDPB), 2020, p. 13). Similarly, the healthcare industry will likely face AI challenges regarding liability when deploying care robots supported by AI systems, when being threatened by cyber-attacks (e.g., data breaches), putting patients' right to privacy at risk (Stephenson and Acklam, 2019, p. 282; Hage, 2017, pp. 255–271). These challenges affect, for instance, the right to respect for private life outlined by Art. 8 of the European Convention on Human Rights (ECHR). These issues not only affect individuals, but also societies, particularly where fundamental rights are at stake (Johansson-Pajala and Gustafsson, 2020, p. 170).[2] Thus, who is liable: the developer or the deployer? (Holzinger, 2016, pp. 119–131).

Lastly, the perspective that AI systems may be fully autonomous may lead to cunning legal arguments to escape developers' and deployers' responsibility (and liability), thereby shifting responsibility to AI systems that lack the legal personality to face accountability (Panezi, 2021, pp. 18–19). Therefore, we argue that, in the course of AI regulation, AI systems should not be viewed as machines acting independently. Rather, in order to prevent faulty AI systems, it is necessary to take a closer look at human participation in this complex ecosystem, which may offer, for now, appropriate accountability solutions (Maclure, 2020, p. 4). In the following section, we examine some key features of accountable AI revealed under the AI Act framework, and discuss whether they may be sufficient to adequately protect fundamental rights.

3. The approach of the EU AI Act to accountable AI

Before examining the AI Act's approach to regulating or introducing accountability measures to protect fundamental rights against harmful AI,

[2] For further reading on the duty of governments to protect citizens' fundamental rights, see Barkhuysen and Van Emmerik (2019).

it may be helpful to briefly revise the different definitions or conceptualisations of accountability.

Accountability may have different meanings or interpretations across different jurisdictions and fields (Bovens, 2010, p. 949). Legal scholars may interpret accountability as responsibility, answerability, or liability (Docksey and Propp, 2023, pp. 2–3), while ethicists may frame it as a moral obligation of private and public organisations to provide an account for their actions (van de Poel et al, 2012, pp. 3–4). Moreover, accountability applied to public administration may be regarded as the government's (and its employees') obligation to exhibit high standards in public service (Newberry, 2015, p. 371). The AI Act itself does not go on to define or conceptualise accountability. However, it does acknowledge the conceptualisation of accountability found in the "Ethics Guidelines for Trustworthy AI" proposed by the High-Level Expert Group (HLEG) (Recital 27 AI Act). The HLEG establishes that accountability requires mechanisms to ensure responsibility for the outcomes of AI systems, both before and after development and deployment (European Commission, 2019, pp. 2, 19). Similarly, the OECD Council on AI has established that accountability in AI regulates the behaviour of actors to develop and deploy AI systems that fully comply with respect for fundamental rights (OECD, 2024, p. 5). Thus, we may argue that the view of accountability, not expressly stated but endorsed by the AI Act, is that accountability relates to the responsibility of developers and deployers to introduce AI systems into the European market that are not contrary to human dignity. This view of accountability is also close to the perspective of legal scholars who regard accountability as the legal responsibility of actors. We may take the opportunity to propose that accountability is essential in society to ensure actors' ownership of their actions. In a societal setting governed by the rule of law, accountability must apply to all actors without exceptions (Constantino and Wagner, 2024, p. 3).

Accountability mechanisms contemplated by the AI Act may include, for instance, introducing human agency and oversight binding requirements, where AI systems are developed and deployed as tools which respect human dignity (Recital 27 AI Act). This approach allows us to infer that the emphasis on providing accountable AI systems is on the human factor to develop and deploy AI systems aligned with human dignity (which, for example, respect fundamental rights). Accountable AI requires developers to build or place AI systems that can be appropriately controlled and overseen by humans (the deployers) (Recital 27 AI Act). Thus, it takes two to tango: the developer to provide functioning AI systems and the deployer (user) to be able to conduct meaningful oversight by controlling or assessing the system and reporting malfunctions (Verdiesen, Santoni de Sio and Dignum, 2021, pp. 143–150, 159). At this stage, it may be worth highlighting that the scope of the AI Act applies to (or is binding on) providers, importers, manufacturers, and deployers (or users) of AI systems used in the EU (Art. 2 AI Act). For the purpose of our analysis, we group developers, importers, and manufacturers under the same category (i.e., developers), and categorise deployers as those organisations or persons that use AI systems for different tasks (e.g., public or private services).

When analysing the human factor in the discussion of accountable AI systems, we may think of humans from two different perspectives. The first relates to human responsibility as a developer of AI systems, considering that AI systems need human intervention as they cannot program themselves or emerge independently (MacKay, 2003). Hence, one may think of AI designers' obligation, or responsibility, to require them to develop products that are not harmful to fundamental rights. A second perspective is the human responsibility as a deployer of AI systems tasked with oversight duties during the deployment of AI systems to prevent or minimise their harmful outputs. This would mean that, in practice, or at least until a court case appears, human oversight responsibilities require human deployers to undertake effective continuous oversight to question and override wrongful AI outputs.

Accordingly, we note that the AI Act has implemented some binding requirements to foster an environment of accountability among the actors involved (developers and deployers). For instance, these requirements may compel developers to follow a risk-based approach to AI systems, where such systems could be categorised into prohibited tools (i.e., those which should not be brought to market), high-risk AI systems, and AI systems with limited risk to fundamental rights and European values (Hanif et al, 2023, pp. 353–354). Some other legislative measures that may promote accountability are the requirements of technical documentation (Art. 11 AI Act), record-keeping (Art. 12 AI Act), accuracy and robustness, and cybersecurity obligations (Art. 15 AI Act). Turning to the binding obligations of AI developers, we can see that, for example, Art. 15 of the AI Act seeks to promote robust AI systems to mitigate risks against citizens' health or other fundamental rights (e.g., data and privacy protection) (Recitals 59 and 75 AI Act). Perhaps the term "robustness", as used by the Act, also refers to accurate AI systems proven to be resilient against cyberattacks (cybersecurity). The robustness of AI systems may also include appropriate datasets and non-bias (OECD, 2024, p. 9). Thus, we understand that Art. 15 interprets robustness as the system's resilience against cyberattacks and ability to provide accurate results, thus preventing errors, faults, or biased outputs that ultimately affect natural persons (Constantino, 2024, p. 404). We may interpret Art. 15 as an attempt to promote a playfield of accountability in innovation, at least binding on deployers (Mahler, 2021, p. 259; Novelli, Taddeo and Floridi, 2022, p. 9). However, there is still much to be seen in practice about the effectiveness (and consequences) of imposing these technical requirements when developing AI systems (Cooper et al, 2022, p. 864). The AI Act has left some gaps or unregulated areas where accountability is crucial. For instance, the Act has not regulated the development and deployment of AI in the intelligence, security, or defence sectors (Constantino and van der Linden, 2024, p. 1), thereby leaving room for different interpretations and standards regarding accountable practices regarding AI systems in these sectors and their effects on society.

The current literature has paid insufficient attention to the duties or responsibilities of deployers of AI systems under the AI Act – particularly the role and qualities of human oversight. In the following paragraphs, we dedicate some time to this matter. For instance, it is thought that Art. 14 only applies to developers of AI systems (Wachter, 2024, pp. 682–683; Demircan, 2023). However, what would be the purpose of introducing human oversight requirements only for developers of AI systems and exempting deployers? In this analysis, we argue that Art. 14 on human oversight obligations does – or, at least, should – apply to both developers and deployers of high-risk AI systems (Koivisto, Koulu and Larsson, 2024, pp. 14–19).[3] Thus, Art. 14 can be read in conjunction with the human oversight obligations outlined in Art. 26(2). For the purpose of our argument, it may be appropriate to read the wording of Art. 14 of the AI Act:

Article 14

Human oversight

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can

[3] Please note that Art. 14 obligations are connected to high-risk AI systems. Thus, the landscape for other AI systems not considered high-risk is not governed by human oversight obligations per Art. 14.

be effectively overseen by natural persons during the period in which they are in use.

2. Human oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular where such risks persist despite the application of other requirements set out in this Section.
3. The oversight measures shall be commensurate with the risks, level of autonomy and context of use of the high-risk AI system, and shall be ensured through either one or both of the following types of measures:
   -(a) measures identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service; (b) measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the deployer.
4. For the purpose of implementing paragraphs 1, 2 and 3, the high-risk AI system shall be provided to the deployer in such a way that natural persons to whom human oversight is assigned are enabled, as appropriate and proportionate:
   (a) to properly understand the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;
   (b) to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;
   (c) to correctly interpret the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available;
   (d) to decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system;
   (e) to intervene in the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure that allows the system to come to a halt in a safe state.
5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 of this Article shall be such as to ensure that, in addition, no action or decision is taken by the deployer

> on the basis of the identification resulting from the system unless that identification has been separately verified and confirmed by at least two natural persons with the necessary competence, training and authority. The requirement for a separate verification by at least two natural persons shall not apply to high-risk AI systems used for the purposes of law enforcement, migration, border control or asylum, where Union or national law considers the application of this requirement to be disproportionate.

The wording of Section 1 of Art. 14 is straightforward. It requires developers to design AI systems that allow human intervention. We may agree that this piece of legislation effectively compels designers to develop tools or processes to allow deployers to conduct effective human oversight to avoid harmful AI that may jeopardise fundamental rights (European Commission, 2019, p. 4). Interestingly, this section refers to natural persons (humans in the loop) to effectively oversee AI systems during deployment. Thus, in principle, Art. 14(1) targets deployers (or designers) of AI systems. However, human oversight requires two actors in this equation in order to have effective human oversight. It is worth noting that the EU legislator is unclear about what "effective" oversight by natural persons means or what responsibilities or actions humans in the loop need to take to make human oversight effective (See Art. 14(1) of the AI Act). Nonetheless, human oversight responsibilities cannot be charged or tasked to one actor, otherwise, it would be pointless to require AI systems built with human oversight interface capabilities but not having actual humans tasked to execute or operationalise them. The previous statement may be supported by the wording of Art. 26(2), which sets an obligation on deployers of AI systems to "assign human oversight to natural persons".[4] Moving forward, Art. 14(2) establishes that the aim of having humans in the loop is to "prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system [are in] used… under conditions of reasonably foreseeable misuse" (Art. 14(2)). The wording provided by the legislator is quite interesting. Firstly, it establishes that "humans in the loop" are there to minimise or prevent the possible harms of high-risk AI systems. Art. 14(2) does not say that AI systems should be built with *self-human-oversight* capabilities to minimise or prevent risks to health, safety, or fundamental rights. Instead, it says that humans have the responsibility to exercise such

[4] See Art. 26(2): "Deployers shall assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support".

control. Secondly, the article chooses an intriguing phrase, “reasonably foreseeable”, which refers to a doctrine that has been primarily applied to the duty of humans (particularly in tort law) to foresee potential risks (Leiman, 2021, p. 252).

Thus, it is unlikely that an AI system with no legal personality or that is incapable of “thinking” outside the box will be tasked with reasonableness and foreseeability (Kowert, 2017, pp. 182–185; Leiman, 2021, pp. 251–253). Hence, it appears that this piece of legislative instrument, at least, paves the way to ascertain the responsibility of deployers to conduct or engage with human oversight. Of course, we are of the view that the framework for human oversight responsibilities established in Art. 14 is to be read in conjunction with Art. 26(2). As the AI Act is very new legislation, there is still room to test Art. 14(2) in court to argue that it provides legal scope to require deployers (users) of AI systems to oversee AI systems to avoid risks to health, safety, and fundamental rights. Art. 14(3) is straightforward and outlines developers’ responsibilities to build AI systems that can allow human-machine interface tools to support human oversight or enable deployers to fulfil their human oversight duties. It may be worth questioning what would happen if a deployer could not conduct human oversight due to the system not having been designed or developed with such technical measures. Then, it is plausible that, under Art. 14(3), deployers may claim non-responsibility for operationalising human oversight obligations.

To complicate the fulfilment of human oversight to foster accountable AI, Art. 14(4) is being drafted almost like a spaghetti. This piece of legislation outlines that human oversight is assigned to natural persons deploying high-risk systems; however, this task (which includes preventing or minimising risks to health, safety, or fundamental rights) is subject to the developer’s ability to build high-risk AI systems that enable such natural persons to conduct human oversight. Art. 14(4) almost implies that developers are solely responsible for enabling or allowing compliance with human oversight duties. For instance, Art. 14(4) establishes that understanding the limitations of the high-risk AI and being able to duly monitor them (lit a), remain aware of overreliance (automation bias) (lit b), decide whether to use, disregard, or question high-risk AI system’s outputs, would depend on how said systems are built (lit d). The binding obligations set out in Art. 14(4) are, arguably, contradictory to Art. 4, which clearly establishes that it is the responsibility of both “*providers and deployers* of AI systems [to] take measures to ensure, to their best extent, a sufficient level of *AI literacy* of their *staff and other persons* dealing with the operation and use [or deployment] of AI systems on their behalf, taking into account their technical knowledge, experience, *education and training* and the context the AI systems are to be used [deployed] in, and considering the persons or groups of persons on whom the AI systems are to be used [deployed]." Thus, it may be appropriate to remind deployers and developers of their obligations, at least under Art. 4, to compel them to employ humans (developers and deployers) with a minimum level of AI literacy (e.g., understanding the ins and outs of algorithmic behaviour) to enable effective human oversight (Neumann, Guirguis and Steiner, 2022, p. 5). The reasoning behind enforcing AI literacy requirements is to have developers and deployers aware of AI capabilities and flaws so they can take appropriate human oversight measures that satisfy an environment of accountability (Green, 2022, pp. 1–3; see also Recitals 20 and 91 AI Act). Lastly, Art. 14(5) also emphasises the requirement of having (at least two) natural persons with the necessary "competence, training and authority" (Article 14(5) AI Act), to conduct oversight in cases of high-risk AI systems outlined in Annex III, point 1(a). This final piece of Art. 14 would allow us to argue that deployers are responsible for including natural persons as part of the human oversight framework. Strangely enough, Art. 14(5) does not apply to high-risk AI systems used for the purposes of law enforcement, migration, border control, or asylum.

To conclude, it may be fair to state that applying Art. 14 of the AI Act will present accountability challenges, such as at what stage and how humans in the loop (deployers) are to intervene or conduct oversight to prevent undesirable AI outputs (Constantino, 2022, p. 12). There is still uncertainty regarding the scope of human oversight for both developers and deployers. Blame shifting may arise and perhaps result in there being too many actors involved in the AI chain, leading to accountability loopholes or gaps (Van de Poel et al, 2012, p. 50). However, fostering AI awareness or education among deployers may provide positive steps toward effective human oversight. AI awareness promotes having more skilled humans who can be prepared to question the AI system, humans who can divert from AI outputs, even in cases where a developer fails or forgets to add technical measures to foster human oversight. Thus, the developer and deployer are responsible for enabling or fostering AI literacy that contributes to effective human oversight. There are no straightforward answers about the *perfect* solution to accountable AI. However, to alleviate current accountability loopholes, promoting and adopting a culture of accountability may be welcomed where the different actors involved in the AI chain can hold each other accountable for their actions (Wagner, de Gooyert and Veeneman, 2023, p. 6). We should also welcome continuous independent human oversight that focuses not on blaming other humans for the faults of AI systems, but rather on an approach that educates others on the acceptable practices regarding the development and deployment of AI systems (Constantino and Wagner, 2024, pp. 8, 14–15). These reasonable approaches to accountability can provide a strong way forward to protect fundamental rights. Lastly, in industries or organisations where the AI Act is not enforceable, other regulations, such as national and international frameworks, can be applied to protect citizens' fundamental rights (Linden, 2021, pp. 5–6). Thus, the absence of regulation should not be an excuse for those willing actors interested in accountability principles.

4. Conclusion

In this Chapter, we have argued that AI systems, at least for now, cannot emerge without human intervention. Thus, we must focus on regulating humans as developers and deployers of AI systems instead of shifting the discussion onto the responsibility of AI systems as if they were fully autonomous beings or capable of legal personality.

The experiences from the last decade have left us with various lessons, such as evaluating AI's effects (negative and positive) on society. AI can be very useful in providing faster and more efficient services to humans, but it can also cause lethal outcomes. For example, while they can help fight crime, it is also clear that AI systems can threaten fundamental rights when they are wrongly or poorly designed and deployed in our societies. Thus, AI systems can discriminate, target vulnerable people, and even breach our privacy. To solve these dilemmas affecting European citizen's fundamental rights, the EU AI Act promotes a framework where developers and deployers of AI systems are charged with certain obligations to close accountability gaps, such as imposing technical requirements onto deployers of AI systems to consider technical documentation, accuracy and robustness, and cybersecurity obligations. At the human level, the AI Act has also considered including human oversight as part of the framework that allows accountability. Human oversight is covered preliminarily in Arts. 14 and 26. However, it is currently being disputed whether, for instance, Art. 14 (human oversight) only regulates developers and exempts deployers from human oversight obligations. We have argued that it takes two (to tango) for accountable AI, meaning that Arts. 14 and 26 (human oversight) should be read together when studying and arguing for the responsibility of both "humans in the loop" (developers and deployers). Developers are responsible for enabling human oversight measures to be incorporated into their AI systems, and deployers are responsible for conducting effective human oversight when they or their organisations use an AI system. This approach can enable effective accountability, promoting citizens' trust when interacting with AI systems (Van Kolfschooten and Shachar, 2023, pp. 1–3; Ng et al, 2020, pp. 7–12). It is hoped that such measures as technical and human requirements will foster accountability among developers and deployers of AI systems, requiring them to introduce AI systems into the European market that are not harmful to humans (Cooper et al, 2022). For instance, rather than blaming computers for their outputs, humans in the loop will be required to move towards a more meaningful human oversight to prevent faulty AI systems and offer explanations to citizens.

Accountable AI may translate as developers' and deployers' joint moral and legal responsibility to allow non-harmful AI in the market. Thus, accountable AI will not be achieved only by adding algorithmic design requirements on developers or designers. Accountability also requires skilled AI deployers to oversee AI systems effectively. Whether the EU AI Act would have positive effects or provide real measures to protect fundamental rights remains to be seen.

References

Appelman, N., Ó Fathaigh, R. and van Hoboken, J. (2021) 'Social welfare, risk profiling and fundamental rights: the case of SyRI in the Netherlands'. *SSRN* [Online]. Available at: https://papers.ssrn.com/abstract=3984935 (Accessed: 18 June 2024).

Barzashka, I. (2023) 'Seeking strategic advantage: the potential of combining artificial intelligence and human-centred wargaming', *The RUSI Journal*, 168(7), pp. 26–32.

Bovens, M. (2010) 'Two concepts of accountability: accountability as a virtue and as a mechanism', *West European Politics*, 33(5), pp. 946–967.

Chander, A. (2017) 'The racist algorithm?', *Michigan Law Review*, 115(6), pp. 1023–1045.

Clifford, D. (2017) 'Citizen-consumers in a personalised galaxy: emotion influenced decision-making, a true path to the dark side?' *SSRN* [Online]. Available at: https://doi.org/10.2139/ssrn.3037425 (Accessed: 3 February 2025).

Constantino, J. (2022) 'Exploring Article 14 of the EU AI Proposal: human in the loop challenges when overseeing high-risk ai systems in public service organisations', *Amsterdam Law Forum*, 14(3), pp. 1-17.

Constantino, J. (2024) 'Article 15 accuracy, robustness and cybersecurity (Forthcoming)', in. *Wolters Kluwer*.

Constantino, J. and van der Linden, T. (2024) 'AI Applications not covered by the AI Act (Forthcoming)', in *Springer*.

Constantino, J. and Wagner, B. (2024) 'Accountability and oversight in the Dutch intelligence and security domains in the digital age', *Frontiers in Political Science*, 6 [Online]. Available at: https://doi.org/10.3389/fpos.2024.1383026 (Accessed: 4 February 2025).

Cooper, A.F., Moss, E., Laufer, B. and Nissenbaum, H. (2022) 'Accountability in an algorithmic society: relationality, responsibility, and robustness in machine learning', in *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 864–876.

Demircan, M. (2023) *Deployers of High-Risk AI Systems: What Will Be Your Obligations Under the EU AI Act?* [Online]. Available at: https://competitionlawblog.kluwercompetitionlaw.com/2023/06/02/deployers-of-high-risk-ai-systems-what-will-be-your-obligations-under-the-eu-ai-act/ (Accessed: 3 February 2025).

DK (2023) *Simply artificial intelligence*. DK Publications. Available at: https://www.dk.com/us/book/9780744076820-simply-artificial-intelligence/ (Accessed: 4 February 2025).

Docksey, C. and Propp, K. (2023) 'Government access to personal data and transnational interoperability: an accountability perspective', *Oslo Law Review*, 10(1), pp. 1–34.

Dougherty, C. (2015) *Google Photos mistakenly labels black people 'gorillas'*. Bits Blog [Online]. Available at: https://archive.nytimes.com/bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas/ (Accessed: 30 September 2024).

Drukarch, H., Calleja, C. and Fosch Villaronga, E. (2023) 'An iterative regulatory process for robot governance', *Data & Policy*, 5 (e8) [Online]. Available at: https://doi.org/10.1017/dap.2023.3 (Accessed: 4 February 2025).

Edwards, L., Schafer, B. and Harbinja, E. (2020) *Future law: emerging technology, regulation and ethics* [Online]. Available at: https://edinburghuniversitypress.com/book-future-law.html (Accessed: 18 June 2024).

Eligon, J. and Williams, T. (2015) 'Police program aims to pinpoint those most likely to commit crimes', *The New York Times*, 25 September [Online]. Available at: https://www.nytimes.com/2015/09/25/us/police-program-aims-to-pinpoint-those-most-likely-to-commit-crimes.html (Accessed: 18 June 2024).

European Commission (2019) *Ethics guidelines for trustworthy AI: High-Level Expert Group on artificial intelligence* [Online]. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (Accessed: 11 June 2024).

European Data Protection Board (EDPB) (2020) *Guidelines 3/2019 on processing of personal data through video devices*. European Data Protection Board [Online]. Available at: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-32019-processing-personal-data-through-video_en (Accessed: 18 June 2024).

Green, B. (2022) 'The flaws of policies requiring human oversight of government algorithms', *Computer Law & Security Review*, 45, 105681 [Online]. Available at: https://doi.org/10.1016/j.clsr.2022.105681 (Accessed: 4 February 2025).

Hacker, P. (2018) 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law', *Common Market Law Review*, 55(4) [Online]. Available at: https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2018095.pdf (Accessed: 11 June 2024).

Hage, J. (2017) 'Theoretical foundations for the responsibility of autonomous agents', *Artificial Intelligence and Law*, 25(3), pp. 255–271.

Hanif, H. et al. (2023) 'Tough decisions? Supporting system classification according to the AI' in Sileno, G., Spanakis, J. and Van Dijck, G. (eds.) *Frontiers in Artificial Intellignece and Applications. Volume 379: Legal Knowledge and Information Systems -.* , pp. 353–358 [Online]. Available at: https://doi.org/10.3233/FAIA230987 (Accessed: 4 February 2025).

Henao, F. (2021) *Why data handling may put a bump on the road to autonomous driving*. Automotive News Europe [Online]. Available at: https://europe.autonews.com/guest-columnist/why-data-handling-may-put-bump-road-autonomous-driving (Accessed: 18 June 2024).

Hill, R.K. (2016) 'What an algorithm is', *Philosophy & Technology*, 29(1), pp. 35–59.

Holzinger, A. (2016) 'Interactive machine learning for health informatics: when do we need the human-in-the-loop?', *Brain Informatics*, 3(2), pp. 119–131.

Human Rights Law Centre (2021) *The Federal Court approves a $112 million settlement for the failures of the Robodebt system*. Human Rights Law Centre [Online]. Available at: https://www.hrlc.org.au/human-rights-case-summaries/2021/9/30/the-federal-court-approves-a-112-million-settlement-for-the-failures-of-the-robodebt-system (Accessed: 18 June 2024).

Johansson-Pajala, R.-M. and Gustafsson, C. (2020) 'Significant challenges when introducing care robots in Swedish elder care' *Disability and Rehabilitation: Assistive Technology*, 17(2), pp. 166–176.

Kang, H.S., Makimoto, K., Konno, R. and Koh, I. S. (2020) 'Review of outcome measures in PARO robot intervention studies for dementia care', *Geriatric Nursing*, 41(3), pp. 207–214.

Kelly, P.A. et al. (2021) 'The effect of PARO robotic seals for hospitalized patients with dementia: a feasibility study', *Geriatric Nursing*, 42(1), pp. 37–45.

Köchling, A. and Wehner, M.C. (2020) 'Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development', *Business Research*, 13(3), pp. 795–848.

Koivisto, I., Koulu, R. and Larsson, S. (2024) 'User accounts: how technological concepts permeate public law through the EU's AI Act', *Maastricht Journal of European and Comparative Law*, 31(3), [Online]. Available at: https://doi.org/10.1177/1023263X241248469 (Accessed: 4 February 2025).

Van Kolfschooten, H. and Shachar, C. (2023) 'The Council of Europe's AI Convention (2023–2024): promises and pitfalls for health protection', *Health Policy*, 138, 104935 [Online]. Available at: https://doi.org/10.1016/j.healthpol.2023.104935 (Accessed: 4 February 2025).

Kowalski, R. (1979) 'Algorithm = logic + control', *Communications of the ACM*, 22(7), pp. 424–436.

Kowert, W. (2017) 'The foreseeability of human–artificial intelligence interactions', *Texas Law Review*, 96, pp. 181-204.

Kuziemski, M. and Misuraca, G. (2020) 'AI governance in the public sector: three tales from the frontiers of automated decision-making in democratic settings', *Telecommunications Policy*, 44(6) [Online]. Available at: https://doi.org/10.1016/j.telpol.2020.101976 (Accessed: 4 February 2025).

Le Sueur, A. (2015) 'Robot government: automated decision-making and its implications for parliament'. *SSRN* [Online]. Available at: https://papers.ssrn.com/abstract=2668201 (Accessed: 18 June 2024).

Leiman, T. (2021) 'Law and tech collide: foreseeability, reasonableness and advanced driver assistance systems', *Policy and Society*, 40(2), pp. 250–271.

Lennox, J. (2020) *2084: artificial intelligence and the future of humanity*. Chicago: Zondervan Reflective [Online]. Available at: https://www.johnlennox.org/shop/24/2084-artificial-intelligence-and-the (Accessed: 6 June 2023).

MacKay, D.J.C. (2003) *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.

Maclure, J. (2020) 'The new AI spring: a deflationary view', *AI & Society*, 35(3), pp. 747–750.

Mahler, T. (2021) 'Between risk management and proportionality: the risk-based approach in the EU's Artificial Intelligence Act Proposal'. *SSRN* [Online]. Available at: https://papers.ssrn.com/abstract=4001444 (Accessed: 18 June 2024).

Muthukrishnan, N. et al. (2020) 'Brief history of artificial intelligence', *Neuroimaging Clinics of North America*, 30(4), pp. 393–399.

Neumann, O., Guirguis, K. and Steiner, R. (2022) 'Exploring artificial intelligence adoption in public organizations: a comparative case study', *Public Management Review*, 26(1), pp. 114–141.

Newberry, S. (2015) 'Public sector accounting: shifting concepts of accountability', *Public Money & Management*, 35(5), pp. 371–376.

Ng, Y.-F., O'Sullivan, M., Paterson, M. and Witzleb, N. (2020) 'Revitalising public law in a technological era: rights, transparency and administrative justice'. *SSRN* [Online]. Available at: https://papers.ssrn.com/abstract=3689497 (Accessed: 18 June 2024).

'NJCM et al v. The Dutch State', ECLI:NL:RBDHA:2020:1878, Rechtbank Den Haag, C-09-550982-HA ZA 18-388 (2020) [Online]. Available at: https://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:RBDHA:2020:1878 (Accessed: 18 June 2024).

Novelli, C., Taddeo, M. and Floridi, L. (2022) 'Accountability in artificial intelligence: what it is and how it works'. *SSRN* [Online]. Available at: https://doi.org/10.2139/ssrn.4180366.

OECD (2024) *Recommendation of the Council on artificial intelligence, OECD/LEGAL/0449* [Online]. Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (Accessed: 18 June 2024).

Panezi, A. (2021) 'Liability rules for AI-facilitated wrongs: an ecosystem approach to manage risk and uncertainty'. *SSRN* [Online]. Available at: https://doi.org/10.2139/ssrn.3768779 (Accessed: 4 February 2025).

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

Ray, P. (2016) 'Internet of robotic things: concept, technologies, and challenges', *IEEE Journals & Magazine* [Online]. Available at: https://ieeexplore.ieee.org/abstract/document/7805273 (Accessed: 30 September 2024).

'Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)' (2024) *Official Journal* L, 2024/1689, 12 July [Online]. Available at: http://data.europa.eu/eli/reg/2024/1689/oj (Accessed: 5 February 2025).

Richardson, R., Schultz, J. and Crawford, K. (2019) 'Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice', *SSRN* [Online]. Available at: https://papers.ssrn.com/abstract=3333423 (Accessed: 18 June 2024).

Rosenblat, A., Kneese, T. and Boyd, D. (2014) 'Workplace surveillance'. *SSRN* [Online]. Available at: https://doi.org/10.2139/ssrn.2536605 (Accessed: 4 February 2025).

Simmons, A.B. and Chappell, S.G. (1988) 'Artificial intelligence-definition and practice', *IEEE Journal of Oceanic Engineering*, 13(2), pp. 14–42.

Sipola, T., Alatalo, J., Wolfmayr, M. and Kokkonen, T. (eds.) (2024) *Artificial intelligence for security: Enhancing protection in a changing world*. Cham: Springer Nature Switzerland.

Stephenson, J. and Acklam, C. (2019) 'Artificial intelligence in care: where does responsibility lie?', *Nursing and Residential Care*, 21(5), pp. 281–283.

Sykes, C.J. (2000) *Big brother in the workplace*Hoover Institution [Online]. Available at: https://www.hoover.org/research/big-brother-workplace (Accessed: 18 June 2024).

Van der Linden, T. (2021) 'Regulating artificial intelligence: please apply existing regulation', *Amsterdam Law Forum*', 13(3), pp. 3-9. Available at: https://doi.org/10.37974/ALF.432 .

Van de Poel, I.R. et al. (2012) 'The problem of many hands: climate change as an example', *Science and Engineering Ethics*, 18(1), pp. 49–67.

Verdiesen, I., Santoni de Sio, F. and Dignum, V. (2021) 'Accountability and control over autonomous weapon systems: a framework for comprehensive human oversight', *Minds and Machines*, 31(1), pp. 137–163.

Wachter, S. (2024) 'Limitations and loopholes in the EU AI Act and AI Liability Directives: what this means for the European Union, the United States, and beyond', *Yale Journal of Law and Technology*, 26(3), pp. 671–718.

Wagner, B., de Gooyert, V. and Veeneman, W. (2023) 'Sustainable development goals as accountability mechanism? A case study of Dutch infrastructure agencies', *Journal of Responsible Technology*, 14, 100058 [Online]. Available at: https://doi.org/10.1016/j.jrt.2023.100058 (Accessed: 5 February 2025).

Whiteford, P. (2021) 'Debt by design: the anatomy of a social policy fiasco – or was it something worse?', *Australian Journal of Public Administration*, 80(2), pp. 340–360.

Winick, E. (2018) *Amazon ditched AI recruitment software because it was biased against women*. MIT Technology Review [Online]. Available at: https://www.technologyreview.com/2018/10/10/139858/amazon-ditched-ai-recruitment-software-because-it-was-biased-against-women/ (Accessed: 18 June 2024).

Wisman, T. (2020) *The SyRI victory: holding profiling practices to account*. Digital Freedom Fund [Online]. Available at: https://digitalfreedomfund.org/the-syri-victory-holding-government-profiling-to-account/ (Accessed: 18 June 2024).