Responsible AI: Law and Advancing Moral Responsibilization

Ana Paula Gonzalez Torres [1]

[1] *Aalto University, Konemiehentie 2, 02150, Espoo, Finland*

Abstract

This paper explores the concept of responsible artificial intelligence (AI) as two-dimensional, related to process and outcome. It discusses the limits of the law and liability to further ethical principles, especially given the AI lifecycle and value chain. Thus, the work relies on concepts of responsible innovation and responsibility as a virtue to discuss individual moral responsibilization narratives. The view is that it requires aiding tools like flagging words to alert people in the AI value chain of the need to implement ethical considerations in specific stages and lead by their particular roles. The paper aims to contribute to the intersection between legal, ethical and technical considerations, leaning on expanding the array of tools to produce responsible AI.

Keywords

Responsible AI, AI governance, Ethical Principles

1. Introduction

Artificial intelligence (AI) is a scientific field with over seventy years under its name [1]. Over the years, it has accomplished triumphs in image and voice recognition [2], even introducing mass availability of large language models [3]. Nonetheless, it has also gathered public outrage because of its ability to reproduce gender [4], racial [5], and transphobic biases[1] even in public services all over the world [6, 7]. Thus, even though it has been around for several decades, recently has emerged a new era within this computer science field which calls for an "ethical", "responsible", or "trustworthy" AI [8]. This new era has seen the publication of multiple principles, guidelines and approaches for the development and adoption of AI-based systems. Followed by a view that society must take responsibility for AI's impact and that individuals such as researchers and developers as well as society should be trained to be aware of their own responsibility when it concerns the development of AI systems with a direct impact on society. Meanwhile, it is up to governments and citizens to determine how systems should be regulated [9].

In Europe, the Commission (EC) installed the High-Level Expert Group on Artificial Intelligence to determine a framework for trustworthy AI. The group published the "Ethical Guidelines for Trustworthy AI" [10], which became the backbone of the first proposal for the regulation of AI systems, known as the "Artificial Intelligence Act" (AI Act) [11]. It was a turning point that witnessed the immersion of ethics and law in a previously strictly computer science field; in such spirit in this paper, 'responsible AI' is to be understood as taking into consideration legal and ethical considerations in the development and adoption of AI-based systems in an effort to make them trustworthy. While principles can emerge as ethical considerations, they can evolve to be part of the law. One example is the transition from ethical guidelines to the regulation of artificial intelligence in Europe. In such case, some but not all, of the ethical principles established in the "Ethics Guidelines for Trustworthy AI" [10] were incorporated in the European Commission's proposed "AI Act" [11]. Acknowledging the difference between ethics and law, the present paper aims to intersect itself with the view that accountability for

Conference on Technology Ethics - Tethics, October 18–19, 2023, Turku, Finland
EMAIL: ana.gonzaleztorres@aalto.fi
ORCID: 0000-0002-2793-7501


[1] https://www.washingtonpost.com/nation/2023/02/07/ai-seinfeld-transphobic-gpt3/

the pursuits of responsible AI should be considered when implementing AI models in real environments [12] and the view that "principles that an organization must follow [..] still need help in practical application" [13]. For instance, while a view sustains accountability as "an action, for which an agent is held responsible" [14], not every ethical consideration is captured by legal provisions.

Thus, in the following, we will examine the limits of the law in fostering "responsible AI" as a two-dimensional concept [14]. One dimension relates to *processes* which have evolved to imply individual accountability of the people making up the AI value chain, leading to *responsibilitization*. The second dimension is the *outcome*, understood as AI-based systems that are produced by following ethical and legal considerations in their development and adoption. It examines how individuals are being asked to lean on their own moral responsibilization to strive for responsible AI. While previous work has mentioned the need for diversity in development teams and education curricula [9], this work aims to contribute with the view that tools that are usually intended to show responsible AI to external parties (e.g., regulators, auditors, stakeholders, etc.) are also to be implemented internally to aid in the responsibilization within organizations. Furthermore, the analysis ventures into liability means to achieve enforceability of rights and ethical principles while ultimately exploring non-legal alternatives. The background driver is the need to discuss the *responsibilization* narrative that "more responsibility and more accountability from the people and organisations involved: for the decisions and actions of the AI applications, and for their *own* decision of using AI in a given application context" [15].

2. Principles, Law & Liability

The "Ethical Guidelines for Trustworthy AI" establish ethical principles which are intended to mirror the EU Charter of fundamental rights in the context of AI systems. For instance, respect for human autonomy, prevention of harm, fairness and explicability. It is said that "AI practitioners should always strive to adhere to them" and try to strike a balance when they appear to be in conflict [10]. Meanwhile, in the transition between ethics and law, we see that the EC proposed "AI Act" has implemented some of the principles by means of requirements for "high-risk" AI systems. For instance, in the ethical guidelines, respect for human autonomy mentioned human oversight, which in the AI Act has been implemented in article 14 and the principle of explicability, which states the need to be transparent, communicate capabilities and purpose, and allow traceability and audibility has been expressed in article 13 and throughout the requirements. On the other side, the principles of prevention of harm and fairness do not seem to have been explicitly considered in the requirements that would be necessary to comply with in order to place in the market or put in service a high-risk AI system *ex* AI Act. As seen, some ethical principles are lost in the transition to regulatory dispositions. It is a necessity as the ethical guidelines themselves are aimed at providing guidance for ethical and robust AI, not lawful AI. Thus, while the principles "offer guidance, they remain abstract ethical principles", and practitioners are "not to be expected to find the right solution based on the principles" [10].

Notice that ethical principles are not enforceable. However, once the AI Act has captured principles by means of regulatory provisions, they could be enforceable as, in case of non-compliance, organisations can be subject to fines and/or liability. In the current legal landscape, the European Union's approach to artificial intelligence involves three initiatives[2]: i) AI Act [11], ii) AI Liability Directive, [16] iii) Product Liability Directive [17], and iv) revision of sectoral safety legislation. The AI Liability Directive covers national liability claims of any person with a view of compensating damage and victims.

Nonetheless, the effects of the AI Liability Directive are circumscribed as the proposal does not harmonise the type of liability for AI systems. Thus, claimants must claim liability based on applicable Union or national rules, in the case of extra-contractual liability, either strict or fault-based negligence liability. Strict liability is the legal responsibility for the damage or loss caused by actions or omissions, regardless of the intentionality of the action, the possibility to control it and the lack of excuse [18]. However, there is usually no compensation for economic loss under a strict liability theory, but only for personal injury or property damage. Such because where there is physical injury and/or property damage, there may be numerous specific torts and grounds of liability that may apply, and it is more

[2] European Commission, A European approach to artificial intelligence, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

likely that compensation will be obtained. In the alternative, negligence liability victims must prove a wrongful action or omission by the entity that caused the damage [18]. It can be established, for example, by demonstrating non-compliance with the provisions of the AI Act or pursuant to other rules set at the Union level[3] [20]. Hence, fault liability has to be established on specific facts related to "high-risk" AI systems that are contrary to AI Act provisions. Based partially on the approach of product liability and the requirements and obligations of the AI Act, some imaginable scenarios [20]:

a) More training with AI is needed, this can result from negligence by training with insufficient data or insufficiently correct data (i.e., data incorrectly labelled or yet to be checked for quality). It would contrast with Article 10, "Quality dataset and data governance" of the AI Act.
b) Incorrect structure of the AI system, this can result from the implementation of neural networks that do not allow sufficiently fine-grained decisions because of insufficient layers or nodes. It would depend on state-of-the-art and existing knowledge but would contrast with article 15, "Accuracy, robustness and cybersecurity", of the AI Act.
c) Insufficient hardware or too slow hardware for the AI, whereby decisions are incorrect or delayed. It would depend on the specific purpose for which the AI is deployed but would contrast with Article 15's requirement of accuracy.
d) Deploying AI for a task and in areas for which it is not suited, this can result from an AI-based system developed and tested in a specific risk that is not in accordance with specific sectorial legislative interventions besides the AI Act.
e) Insufficient precautionary measures, for instance, allowing AI to produce outcomes without human intervention in case of irregularities, would be against Article 14, "Human oversight" of the AI Act.

In such scenarios, errors or insufficient measures can lead to an AI-based system that does not function correctly according to the standards stated in the AI Act and, hence liability. While liability can discourage responsible behaviour and thus contribute to the occurrence of serious accidents rather than to their prevention, [14] the goal to hold parties accountable and do justice to victims can also function as a contributor to responsible AI. The ultimate goal of any liability framework is to provide legal certainty to all parties, whether it be the producer, the operator, the affected person or any other third-party [21]. To apply fault-based liability rules, one must be able to trace harm back to human behaviour [22]. However, the opacity of certain AI systems and complex AI value chains can make it difficult or prohibitively expensive for victims to identify the liable entity and prove the requirements for a successful liability claim [23]. Even though the proposed AI Liability Directive introduces alleviations of the burden of proof, the claimant still needs to establish negligence, which might become a barrier to obtaining compensation and establishing accountability, especially in the case of fundamental rights that underpin legal and ethical principles.

3. The practicalities of responsible AI

In practice, the lifecycle of an AI system involves different parties throughout the different stages. Its lifecycle can be thought of as comprising conceptualisation, data, development, deployment, maintenance, and retirement [24]. The involvement of different parties creates a multistakeholder ecosystem leading to different AI value chain scenarios:

1. Internal AI development and deployment (in-house), in which a single entity develops and deploys a model.
2. One entity develops an AI system for another entity (AI system contracting), the contractor develops and assists a contracting entity in deploying a model.
3. One entity writes the code and trains the system, then sells access through a branded application or API (AIaaS, restricted AI system access), a company deploys a model, the client sends input data and gets output data. In this case, the client pays for access.
4. A vendor writes code for an AI system but does not pre-train it or provide training data to purchasers (software with AI code), a vendor sells software with AI code, and a purchaser adds data to finish an AI system.

[3] Article 4 "Rebuttable presumption of a causal link in the case of fault", AI Liability Directive.

5. Vendors of learning AI systems (AI as a product), a vendor develops an AI system as part of a software, sells software while a purchaser further develops the model with new data.
6. Initial development by one entity and fine-tuning by another (AI system fine-tuning), developer 1 sells its model while developer 2 adds data to fine-tune the AI system.
7. One entity integrates different AI systems into a new one (AI model integration), developer 1 and developer 2 sell models to an integrating developer.

The various AI value chain scenarios signal several challenges as, in multiple cases is not clear which entity will typically be responsible for the AI-based system on a liability basis. While the causal condition may be more easily met when fewer stakeholders are directly involved, causality remains an issue because of the multiplicity of stakeholders involved in the process of producing AI systems that bringing negative social consequences. In this light, regulatory measures, such as the AI Act and accompanying liability regulation, seem unable to intervene in a complex web of organisations' internal structures to determine individual accountability for regulatory compliance nor to enforce or foster ethical principles [22]. The multistakeholder environment seen in the AI value chain signals the need for responsible AI to prevent harm that could induce liability, incentivising individual parties to check the responsible AI practices of their counterparties as they could be joint and severally liable [16]. Liability would ensure that the different stakeholders are aware and interested in their' and others' responsible AI practices through the lifecycle, as the potential of liability provides the incentive to prevent damage from non-compliance.

Thus, the AI Act and AI Liability Directive seem to be limited in their ability to influence organisations' internal processes to produce responsible AI as an outcome. While liability could be an important contributor to responsible AI, the goal of furthering compliance with AI regulation and ethical principles could be aided by alternative measures. For instance, by designing tools for individual moral responsibilization. It would encompass implementing internal practices at different stages and geared towards different individuals, from data scientists, product owners, chief AI officers, corporate social responsibility officers, and development teams to AI adopters or deployers [9]. It is the view of this work that responsible AI would involve a change in perspective focusing on practical *processes* that go inwards, not just outwards, and building towards *outcomes*.

4. Responsible AI as responsible innovation

Responsible innovation is considered to start at already early stages of technological research and development. Thus, innovators anticipate potential uses and societal consequences, risks and benefits of technologies and proactively aim to contribute to ethical principles and societal challenges [14]. This view calls for "responsibility as a virtue", which refers to an individual's inclination to assume or take responsibilities and an awareness of relevant normative demands. In this case, moral responsibilization would rely on such an inclination to appeal to individuals in specific roles of the AI value chain to further not just legal demands but ponder on ethical considerations in their contribution through the process of producing an AI system. To take into consideration process requirements for responsible innovation (e.g., anticipation, reflexivity, inclusiveness and responsiveness) or in terms of products that embed relevant values [14]. It is based on the understanding that, ethically, innovation comes from a willingness to care for others through the lenses of responsible innovation. The desire to take on several more specific responsibilities goes beyond legal requirements or obligations, which stands to permeate the AI value chain to reach individuals. While new normative demands may arise during the innovation process, individuals could be better positioned if they can recognize and respond to such normative demands proactively.

Following the moral *responsibilization* narrative, individuals making up the AI value chain (e.g., data scientists, product owners, chief technology, AI champion, machine learning engineers, adopters of AI-based systems, etc.) if their *process* is poor and the *outcome* faulty, then all the individuals involved in the AI value chain are deemed morally responsible [15, 18]. This type of responsibility falls from ethical principles but relies on moral accountability. Thus, any individual in the AI value chain is fully responsible, morally, for what the whole value chain outputs. This may encourage some or even all individuals to refrain from acting or even abandoning the value chain, while others will try to avoid these outcomes. Thus, the design of proper incentives to encourage agents to take some reasonable and

limited moral accountability is to be pursued. In an ethical context, moral hedging can be done by facilitating a better understanding of an individual's duties towards proactive care of the system affected [18]. Note that while individuals making up the AI value chain at different stages of the AI lifecycle might be better able than end-users to anticipate hazards and guard against their effects, this capability is limited to their functions and roles. A particular consideration, in the case of "general purpose AI" such as ChatGPT and other similar applications, which can be used for a variety of use cases, the foreseeable consequences expand exponentially. They could require a moral responsibilization of even end-users. This consideration is worth noticing but is outside the scope of this paper.

Following the moral *responsibilization* narrative, individual accountability to follow not just the law but ethical principles requires translating values into forward-looking design requirements and tools that meet as many values as simultaneously possible [14]. Mainly because individuals making up the AI value chain are not able to consciously implement ethical considerations as a derivative of their condition as humans but need aiding tools [25]. Such because engineers and computer scientists may see their responsibility as focused on the quality and safety of a particular AI system rather than on large-scale social issues. They may be unaware of the broader set of implications [12]. In addition, there is uncertainty regarding one's scope of moral *responsibilization*; engineers and computer scientists who focus on the development of systems may have limited influence within their organizations. They may expect managers, product owners, legal officers, or corporate social responsibility staff to assess broader social and ethical issues. This view would lead to the 'many hands' problem, where accountability for responsible AI is distributed and disarrayed [12].

Thus, individual *responsibilization* can be seen as a means to instil a sense of accountability for the final outcome and a sense of contributing to a responsible AI that could impact society and their organization's legal obligations. It would require many stakeholders involved in shaping AI to be functionally able to recall a concrete division of labour for specific legal and ethical considerations. If companies fail to resolve these challenges, they may face public scrutiny as well as financial and legal risks and reputational harms [12]. The presented view is that tools previously employed for external validation of responsible AI are to be implemented inwardly. They shall be part of *processes* that touch individuals that make up the AI value chain to aid them in their moral *responsibilization* to contribute towards accomplishing legal and ethical principles. For instance, establishing flagging words according to specific roles and then requiring the people in those roles to follow up their tasks by performing a review of the checklist in the "Ethical Guidelines for Trustworthy AI" or implementing tools that could further those principles. In the presented approach, the "flagged words" signal the existence of relevant issues and incentivize individuals to translate values into tools that could meet legal and ethical principles at each stage of the AI lifecycle [14]. For example, implementing model cards [26], factsheets [27], data nutrition labels [28] or implementing operations that allow for tracing and auditability [29]. In Figure 1, the presented "flagging words" could incentivize pondering principles and relevant issues, which could help individuals like data scientists realize that they do not have enough contextual information and would need to involve, for example, impacted end-users to implement data in a non-stigmatizing manner [30].

Ethical Principles	Stage	Role	Flagged words
Respect for human autonomy	Conceptualization Development Deployment Maintenance	Product owner Software developer AI adopter/deployer	Human oversight; Autonomy of human beings.
Prevention of harm	Development Deployment Maintenance Retirement	Model developer AI adopters/deployer	Malicious use; Vulnerable people; Natural environment.
Fairness	Conceptualization Data Maintenance Retirement	AI adopters/deployer Data scientist Model developer	Unfair bias; Discrimination; Stigmatisation; Redress.
Explicability	Development Deployment Maintenance Retirement	Data scientist Model developer AI adopter/deployer	Auditability; Traceability; Transparency in capability and purpose.

Figure 1: Ethical principles, stages of the AI lifecycle and roles concerned with flagged words

It could be said that there is the possibility of unfair attribution of responsibility if the benchmark is defined in terms of eventual societal outcomes. Still, it would be possible to establish responsible AI innovation by the *outcomes* and innovation *processes* for which we hold innovators accountable or for which innovators can reasonably take responsibility [14]. The goal is not to determine moral *responsibilization* for general societal outcomes of a produced AI system but how their expected *outcomes* included ponderations of ethical principles. While other proposed tools' ethical considerations are usually led by an "ethical AI board" [31], a single person "AI champions" [32] or performing "ethics as a service" [33] in the present paper, the view is that responsible AI considerations are to be part of the internal *process* available to individuals in the AI value chain. Meanwhile, their capability to implement broader considerations is determined by the availability of aiding tools to comply with their moral *responsibilization*.

5. Conclusion

The production of AI-based systems is growing ever-increasingly, and its ramifications are felt everywhere. The present paper discussed the "loss in translation" that happens in the process from ethical principles to AI systems regulation. In such a line, responsible AI is considered two-dimensional, encompassing *process* and *outcome* as touched by law and ethics. In particular, the limits of the law are explained by the narrow scope of liability frameworks to capture ethical principles. "Ethical Guidelines for Trustworthy AI" examination shows how the proposed AI Act has partially captured the ethical principles for AI systems. For instance, human autonomy and explicability as explicitly captured in the requirements for "high-risk" AI systems. Even though prevention of harm and fairness are not explicitly mentioned in the proposed law, notice that ethical principles are meant as guidance and are self-declared "abstract" [10]. While the principles do not touch on the lawfulness of AI systems, understanding the legal landscape surrounding AI systems uncovers the limits of the adjunct AI Liability Directive. While the AI Act requirements and obligations can function as a basis for liability based on faulty compliance. The limited scope of the AI Liability Directive provides an uncertain incentive to lawful or ethical principles for the development of AI systems. It is the view that the obstacles to liability are the need to ascertain that an AI-based system is defective, that it depends on national legislation with their different bases for liability, the context in which it was developed, opaqueness and the multiple stakeholders' AI value chain.

Following the examination of the obstacles to liability, we discuss the practicalities of the AI lifecycle from conceptualization, data, development, deployment, maintenance and retirement, as it involves multiple stakeholders (e.g., data scientists, developers, deployers, product owners, AI adopters) who are causally involved in the process of producing AI systems. It is discussed how such intricate involvement of different parties makes it prohibitive to establish clear responsibility. Consequently, the conclusion is that the AI Act just partially captures ethical principles and that liability could have diverse impacts because of the complex AI value chain that we turn to responsibilization.

Employing the concepts of responsible innovation and responsibility as a virtue to further individual moral *responsibilization* as the resource that could close the gap in the aim to further ethical principles through the lifecycle of AI. The present work accepts that not just because the AI value chain is made of people, they will implement ethical choices consciously. It requires acknowledging that responsibility as a virtue is cultivated, not demanded. Thus, responsible AI, as encompassed in the narratives of moral *responsibilization*, is to rely on tools such as "flagging of words" to be part of the *process* of producing an AI system. The presented flagging words are correlated to specific stages of the AI lifecycle and roles. It is the view that people who make up the AI value chain can be alerted, and by relying on their moral responsibilization, relevant issues can spark further ethical consideration and lead to the conscious implementation of tools. While this view is part of a broader effort to install ethical theories in the scientific field of AI, it shall be the subject of further research. Future work is to examine the viability of such a tool and its impact on fostering responsible AI practices as a *process* and *outcome*.

6. Acknowledgements

This research was possible with the support of the CRAI-CIS Research Group, Aalto University, Department of Computer Science.

7. References

[1] S. J. Russell, P. Norvig, E. Davis, Artificial intelligence: a modern approach, 3rd ed., Prentice Hall series in artificial intelligence. Upper Saddle River: Prentice Hall, New Jersey, 2010.

[2] A. Longo, G. Scorza, Intelligenza artificiale l'impatto sulle nostre vite, diritti e libertà, 1st ed. Milano, 2020.

[3] S. Johnson, N. Iziev, A.I. Is Mastering Language. Should We Trust What It Says?, The New York Times, 2022. version 4, 2023. URL: https://www.nytimes.com/2022/04/15/magazine/ai-language.html.

[4] J. Buolamwini, T. Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, presented at: FAT, Jan. 2018. URL: https://www.semanticscholar.org/paper/Gender-Shades%3A-Intersectional-Accuracy-Disparities-Buolamwini-Gebru/18858cc936947fc96b5c06bbe3c6c2faa5614540.

[5] L. Sweeney, Discrimination in online ad delivery, Commun. ACM, vol. 56, no. 5, pp. 44–54 (2013). doi: 10.1145/2447976.2447990.

[6] J. Angwin, J. Larson, S. Mattu, L. Kirchner, Machine Bias, ProPublica, (2016). version 4, 2023. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[7] F. Chiusi, S. Fischer, N. Kayser-Bril, M. Spielkamp, Automating Society 2020, Algorithm Watch, (2020). URL: https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf.

[8] A. Jobin, M. Ienca, E. Vayena, The global landscape of AI ethics guidelines," Nat. Mach. Intell., vol. 1, no. 9, pp. 389–399 (2019). doi: 10.1038/s42256-019-0088-2.

[9] V. Dignum, Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way, 1st ed., in Artificial Intelligence: Foundations, Theory, and Algorithms. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-30371-6.

[10] Independent High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy AI, (2019). URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[11] European Commission, Proposal for a Regulation of the European Parliament and of the Council. Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and

Amending Certain Union Legislative Acts, Brussels, 21.4.2021, COM(2021) 206 final. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.
[12] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia., S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115 (2020). https://doi.org/10.1016/j.inffus.2019.12.012.
[13] D. Schiff, B. Rakova, A. Ayesh, A. Fanti, M. Lennon, Principles to Practices for Responsible AI: Closing the Gap, 2020. doi: 10.48550/ARXIV.2006.04707.
[14] I. Van De Poel, M. Sand, Varieties of responsibility: two problems of responsible innovation, Synthese, vol. 198, no. S19, pp. 4769–4787 (2021). doi: 10.1007/s11229-018-01951-7.
[15] V. Dignum, Responsible Artificial Intelligence -- from Principles to Practice, 2022. URL: http://arxiv.org/abs/2205.10785
[16] European Commission, Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), Brussels, 28.9.2022, COM(2022) 496 final.
[17] European Commission, Proposal for a Directive of the European Parliament and of the Council on Liability for Defective Products, Brussels, 28.9.2022, COM(2022) 495 final.
[18] L. Floridi, Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., vol. 374, no. 2083, p. 20160112 (2016). doi: 10.1098/rsta.2016.0112.
[19] European Commission, Directorate General for Justice and Consumers, Liability for artificial intelligence and other emerging digital technologies, 2019. URL: https://data.europa.eu/doi/10.2838/573689.
[20] E. T. T. Tai, Liability for AI Decision-Making, in *The Cambridge Handbook of Artificial Intelligence*, L. A. DiMatteo, C. Poncibò, and M. Cannarsa, Eds., 1st ed., Cambridge University Press, 2022, pp. 116–131. doi: 10.1017/9781009072168.014.
[21] European Parliament resolution of 20 October 2020 with recommendation to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)).
[22] B. Schütte, L. Majewski, Private Liability for AI-related Harm – towards more predictable Rules for the Single Market, Mark. Compet. Law Rev., pp. 123-159, (2022). doi: 10.34632/MCLAWREVIEW.2022.10741.
[23] A. Engler, A. Renda, Reconciling the AI Value Chain with the EU's Artificial Intelligence Act, 2022. URL: https://www.ceps.eu/ceps-publications/reconciling-the-ai-value-chain-with-the-eus-artificial-intelligence-act/.
[24] D. De Silva, D. Alahakoon, An artificial intelligence life cycle: From conception to production, Patterns, vol. 3, no. 6, p. 100489 (2022). doi: 10.1016/j.patter.2022.100489.
[25] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, L. Floridi, Operationalising AI ethics: barriers, enablers and next steps, AI Soc., vol. 38, no. 1, pp. 411–423 (2023). doi: 10.1007/s00146-021-01308-8.
[26] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. Raji, T. Gebru, T, Model Cards for Model Reporting, in Proceedings of the Conference on Fairness, Accountability, and Transparency (2019) 220–229. https://doi.org/10.1145/3287560.3287596.
[27] M. Arnold, R. K. E. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilovic, R. Nair, K. N. Ramamurthy, D. Reimer, A. Olteanu, D. Piorkowski, J. Tsay, K. R. Varshney, FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity, 2019. URL: http://arxiv.org/abs/1808.07261.
[28] S. Holland, A. Hosny, S. Newman, J. Joseph, K. Chmielinski, The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards, 2018. URL: http://arxiv.org/abs/1805.03677.
[29] T. Granlund, V. Stirbu, T. Mikkonen, Towards Regulatory-Compliant MLOps: Oravizio's Journey from a Machine Learning Experiment to a Deployed Certified Medical Product, SN Comput. Sci., vol. 2, no. 5, p. 342 (2021). doi: 10.1007/s42979-021-00726-1.
[30] C. D'Ignazio, L. Klein, Data feminism, MIT Press, 2020, ISBN 9780262044004.
[31] D. Neyland, Bearing Account-able Witness to the Ethical Algorithmic System, Sci. Technol. Hum. Values, vol. 41, no. 1, pp. 50–76 (2016). doi: 10.1177/0162243915598056.

[32] Telefonica, Telefónica´s Approach to the Responsible Use of AI, version 4, 2023. URL: https://www.telefonica.com/en/wp-content/uploads/sites/5/2021/08/ia-responsible-governance.pdf.
[33] J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. Mökander, L. Floridi, Ethics as a Service: A Pragmatic Operationalisation of AI Ethics, *Minds Mach.*, vol. 31, no. 2, pp. 239–256 (2021). doi: 10.1007/s11023-021-09563-w.