A Vision on What Explanations of Autonomous Systems are of Interest to Lawyers

DOI:
10.1109/REW57809.2023.00062

Document Version
Accepted author manuscript

A Vision on What Explanations of Autonomous Systems are of Interest to Lawyers

Miriam C. Buiten
Law School
University of St. Gallen
St. Gallen, Switzerland
miriam.buiten@unisg.ch

Louise A. Dennis
Department of Computer Science
University of Manchester
Manchester, UK
louise.dennis@manchester.ac.uk

Maike Schwammberger
Department of Informatics
Karlsruhe Institute of Technology
Karlsruhe, Germany
schwammberger@kit.edu

***Abstract*—As (semi-) autonomous systems become more prevalent, accountability for their actions in legal cases becomes crucial. However, understanding the decision-making process of these complex systems can be challenging. System explainability offers a solution by providing insights into how these systems work. In our research vision, we focus on identifying the types of explanations that lawyers need in litigation involving autonomous systems. With an increase in autonomy, these systems get increasingly complex, with some systems even being "black-" or "grey-box" systems where large amounts of their decision making is obscured. By bridging the gap between technology and the law, we aim to enhance the legal process surrounding autonomous vehicles.**

***Index Terms*—Explainability, accountability, transparency, autonomous systems, artificial intelligence, lawyers**

I. Introduction

Recent incidents involving self-driving cars illustrate the need to hold autonomous systems accountable for the harms that they cause.[1] Civil liability lawsuits play a crucial role in seeking justice for victims. In such cases, judges are increasingly confronted with complex autonomous systems. We focus on autonomous systems in general, meaning systems that make high-level decisions and that may include components based on statistical models or neural networks. Such autonomous systems often operate as "black- or grey-boxes", sometimes in order to protect manufacturing secrets, obscuring large amounts of their decision-making process.

This opacity poses a challenge in tort litigation, where courts have to attribute losses to responsible parties. It makes it difficult to determine how and why complex autonomous systems make decisions, and consequently for victims to identify the liable person and prove the requirements for a successful liability claim.[2]

L. Dennis was supported by EPSRC grants Computational Agent Responsibility (EP/W01081X/1) and TAS Verifiability Node (EP/V026801) and M. Schwammberger was supported through the MWK Baden-Württemberg within the Innovation Campus for Future Mobility.

[1]See e.g. S. Suber and M. Saxon, "First Lawsuit Filed for Tesla Autopilot-Related Death Involving a Pedestrian", www.winston.com/en/product-liability-and-mass-torts-digest/first-lawsuit-filed-for-tesla-autopilot-related-death-involving-a-pedestrian.html.

[2]See e.g. the 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), 28.9.2022 COM(2022) 496 final, at 1.

One solution to address the "black-box" problem is the development of *explainable* software systems. This is also referred to as interpretability and transparency. These approaches aim to provide explanations for how algorithms reach their conclusions or predictions. Deriving explainability requirements [1] requires an interdisciplinary viewpoint: Insights from computer science and different social sciences have to be considered [2].

In this research vision, we focus on requirements that lawyers have for system explanations. We recognise that AI explainability has the potential to assist lawyers in various ways, including policymakers responsible for designing AI regulations.[3] The content, scope, and timing of an explanation vary depending on the specific goal or action it is intended to support [3]–[5].

Our research vision focuses on a specific area where explanations of autonomous systems are needed, namely tort litigation. Courts and attorneys require explanations to verify accident causes and assign responsibility accurately. We explore whether contemporary explainability approaches can provide the necessary information for courts to evaluate tort claims involving autonomous systems. We combine law and computer science to shed light on the opportunities and challenges of using system explanations in civil litigation. Explanations need to align with the legal elements required to succeed in a lawsuit, such as fault, product defect and causality. It is important to understand *how the system was explaining itself or how it was explained to users* and *how the system development process was structured in order to mitigate any unpredictability in the system.*

We focus on the challenges that we perceive for engineering explanations for lawyers, as well as the types of explanations that would be necessary in court. In the next Sect. II, we discuss some background and related work on explainability engineering to explore the state of the art in the field. We subsequently give details on the central topics of *why*, *when*,

[3]See on explainability and transparency e.g. the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts, 21.4.2021 COM(2021) 206 final and the Recommendations of the High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, European Commission, 2018.

and *what* to explain in the case of explanations for lawyers in Sect. III. We conclude with a summary and an outlook in Sect. IV.

II. Explainability Engineering

A *self-explainable software system* is one that provides explanations for its chosen behaviour [6]–[8]. When engineering a self-explainable system, a collection of questions must be kept in mind:

A *"Why explain?"*
B *"When to explain?"* and
C *"What to explain?"*

We focus on these three questions from a technical viewpoint in this section and give answers to these questions from the perspective of lawyers in Sects. III-A to III-C.

A. Why explain?

The process of explaining consists of several phases, where in one of the first phases, the need for an explanation must be identified. Such a need could be some unexpected system behaviour or that an end-user requests an explanation. Only if an explanation is necessary, we need to consider when and what to explain.

The need for an explanation can be identified by monitoring and analysing system behaviour and the environment during run-time, as is sketched for e.g. the modular self-explainability framework MAB-EX [8]. Through such observations, anomalous behaviour that requires an explanation can be found. To reduce the search space for finding such behaviour, approaches exist that classify behaviour with similar reasons [7]. The claim that explanations are needed for anomalous or unexpected behavior was substantiated by [9], who found that explanations were required when a route-following robot deviated from its planned route.

B. When to explain?

For the question about when to explain, the following three phases are often considered [10], [11]:

1) "*A priori*" (explain before an event happens),
2) "*During*" (Explain while an event happens), and
3) "*A posteriori*" (explain after an event happened, "forensic explanation").

An a priori explanation is helpful for cooperative manoeuvres in human-machine interaction; *e.g. before an automated vehicle enters an unknown situation and has to transfer control back to the human driver, it explains the situation to the driver*. Explanations during events can be useful, e.g., in emergency situations; *the automated vehicle has to do an emergency braking manoeuvre, and prepares the passengers for this with a quick explanation*. This is particularly important when something unexpected has occurred. Finally, explanations after an event are of interest, e.g., if an irregularity or even an accident occurred, and it must be determined what caused this event; *e.g. an accident with two automated vehicles occurred and an engineer or court needs to verify whether a malfunction in one of the vehicles caused the accident*.

It is important to note that the time when an explanation is provided also influences the type of explanation; An explanation given well before an event may be more detailed than an explanation during an emergency manoeuvre. A longer *explanation dialogue* between the explainer and the explainee may also be considered (e.g. [12]–[15]). This is of particular value in responding to "why not?" questions [16] where the scope of potential answers can be large and it can be useful to ask follow-up questions to narrow down the explanation to information of relevance to both parties. In critical situations, critical data (such as "health monitoring data") may be requested continuously [9].

C. What to explain

What to explain is strongly linked to the question to whom to explain. We can identify several types of explanation recipients with different requirements regarding the type and depth of an explanation. The IEEE Standard P7001 on Transparency for Autonomous Systems [17], [18] distinguishes users, the general public and bystanders, validation and certification agencies and auditors, incident investigators, and advisors in administrative actions or litigation. In [19], a procedure is explored with which explanations can be derived automatically from system models, for different explainee types. They propose that, for refining explanations towards different explainees, some system details must be hidden, and some additional information might have to be added, e.g. from requirements engineering documents. In terms of content, according to the results in [9], explanations should be framed in terms of a "mental model" of the system. In particular, engineers wanted to understand *what* the system believed and *what it intended* (i.e., was trying to do).

P7001 also emphasises that transparency for those involved in litigation should include documentation of the development process of a system in terms of quality management, audit trails, risk assessments and governance with a particular emphasis on consideration of ethical behaviour.

The question of *what to explain?* also entails the question of *how to explain?*. Much work on explainability focuses on the use of visualisations for understanding statistical models (e.g., [20]) with a particular emphasis on understanding which features of the input were important to some classification. This set of techniques is often broadly grouped under the acronym XAI. The associated tools are generally intended for use by experts with a strong working knowledge of both how the system and the explainability mechanism work. Following [2] there has been a significant strand of work looking at *counterfactuals* as explanations – establishing which inputs would need to change in order to alter the behaviour of the system. However, such techniques currently fail to distinguish satisfactorily between causation and correlation which has led some to caution against over-reliance on the methodology [21]. "Black Box" natural language generation systems such as chatGPT can also be leveraged in producing explanations [22]. However such approaches naturally lead to additional concerns

about the accuracy of the explanation itself, if parts of the generation process have been opaque or statistical.

III. Explanations for Lawyers

A. Why explain to lawyers?

Autonomous systems can play a role in tort liability cases in various context. For instance in case of accidents involving consumer products, autonomous vehicles, or medical systems, or when algorithmic decision-making by companies or government agencies produce biased outcomes [23].

Establishing *accountability* for autonomous systems is crucial when their decisions result in harm [24]. In order to attribute responsibility for the actions and decisions made by autonomous systems, courts need to verify what happened, what caused it to happen and who was responsible for it. By seeking explanations of autonomous systems, courts can gain insight into the factors and processes that influenced the system's behaviour, and identify errors or defects that occurred during their operation.

Attributing responsibility for losses is challenging when systems act or decide independently. The unpredictability of their actions makes it difficult to align with traditional legal concepts such as fault, product defect and causality [25]. For instance, in cases involving accidents with autonomous cars, courts must evaluate whether inadequate design by the producer or the user's lack of attention while driving is to blame [26]. In healthcare, uncertainty arises regarding the accountability of physicians when autonomous systems provide inaccurate recommendations for diagnosis and treatment [27]. When autonomous systems impact individuals' rights and liberties, e.g. when used by credit agencies or law enforcement, courts may demand explanations to ensure *due process and fairness* [23].

B. When to explain to lawyers?

Courts will generally require two types of explanations. Primarily, courts require explanations after the event in order to understand how the system operated and what caused the accident or harm. Such explanations help determine if the AI system adhered to legal standards, identify defects or biases, and assign liability. These explanations need to be comprehensible to individuals who are not experts in autonomous systems. [28], [29].

Secondly, courts may require second-order explanations: they may inquire what explanations were provided to users before and during system operation to ensure safety and effective human-machine interaction.This may assist courts in determining whether a user met their duty to monitor the system. Courts may also need to review information provided about the system to certification bodies and other regulatory agencies, in order to establish whether the autonomous system complied with regulatory standards.

C. What to explain to lawyers?

Explanations of autonomous systems need to allow claimants (or their attorneys) to meet the legal requirements for a tort claim, and allow courts to evaluate these. Claimants must firstly be able to recognise that they are affected by an autonomous system [30]. Courts then need to verify whether the producer or the user is liable. Producer liability requires that the product was defective and the defect caused the harm [26]. User liability generally requires that the user was at fault or breached a duty, and that this breach of duty caused the harm.[4] In both settings, the claimant bears the burden of proof that the product was defective, or the user at fault, respectively.

In a case against a producer, the claimant thus needs explanations allowing her to prove a product defect: e.g. that a sensor failed, hardware was defective or the system was not sufficiently designed or trained for the context in which it operated [26].[5] Courts may also need to know the overall failure rates of the fleet of systems to evaluate in what situations and how often the system malfunctions. Developers may not have the same level of control over automated systems that manufacturers have over the functioning of traditional products [31]. This lack of control raises questions about the expected level of safety for systems that are designed to make autonomous decisions or take actions [25]. The court's assessment of product defect therefore involves determining whether the system fulfilled its promise to function autonomously and was genuinely marketed as an autonomous system [25]. If human machine interaction is required, courts also need information on whether the system adequately instructed the user to intervene at a critical moment.

If the user failed to monitor the system, courts need to verify if the user breached a duty of care and was therefore liable under fault (or negligence) liability. Determining fault becomes ambiguous when system actions cannot be reasonably predicted and users have limited control [31], [32]. To establish the user's duty of care, courts may need to know under what circumstances the user could rely on the system, what information the user had to critically evaluate the system's decisions and whether the user could override or otherwise control the system [25].

Autonomous systems also pose challenges to traditional tort paradigms if the reasoning for their decisions cannot be understood [27]. This reasoning could be relevant to e.g. a physician following a recommendation of an autonomous system, in order to critically evaluate if it is correct. Courts may need explanations of how such a system reached its decision and whether it should have been evident to the physician that the decision or recommendation was erroneous.

To prove causality, courts need to verify whether a breach of duty on the part of the producer or the user materially contributed to the damaging event. In order to do this, courts need to know what happened and what caused it to happen.

[4] In some circumstances, users may be strictly liable, e.g. in several jurisdictions for harm caused by their motorised vehicles. In this case, claimants only need to show causality and harm.

[5] Under EU product liability law, the relevant criterion is whether a product meets consumer expectations. See Council Directive 85/374/EEC on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [1985] OJ L 210/29 (Product Liability Directive), Art. 6.

Explanations of autonomous systems can allow courts to reconstruct what conditions caused an autonomous system to fail [33].

Finally, in some legal systems claimants need to prove that the claimant breached a duty, or a specific legal provision, to establish liability. Courts may require an insight into a broader set of decisions by the system to understand if the harmful decision or action constitutes a breach of duty, right or obligation [25]. The legal standard may require something different from the system being accurate: for instance, if an algorithm takes biased decisions discriminating a particular group, the legally relevant question is not if the underlying dataset was biased, but if the outcome violated the right not to be discriminated against [25].

Overall, explanations in tort litigation need to be aligned with legal requirements. Explanations need to provide information on the set of facts that must be proven to successfully claim damages [34]. Courts may need to know how a system arrived at a certain decision or under what conditions it tends to fail. Explanations need to allow claimants to prove to the court that some characteristic of the system, or some conduct in developing and deploying the system, or some relationship between the two, falls short of the relevant legal standard [30]

An interplay therefore exists between explainability and legal requirements. Explanations of autonomous systems can assist claimants in finding out what happened, while procedural rules can alleviate the burden of proof placed on them to prove what happened. Legal research is actively examining the necessary evolution of civil liability laws to address harms caused by autonomous and self-learning systems [24], [25], [27], [31], [32], [35]–[37]. The EU has proposed an alleviated burden of proof for liability claims involving AI systems,[2] as well as modernised rules for product liability.[6]

To ensure feasible and useful explanations for autonomous systems [4], a continuous exchange between legal scholars and computer scientists is crucial. Courts can stimulate diverse explanations for different legal settings and audiences [23].

IV. Outlook

In this research vision we summarised and pointed out the challenges and opportunities of developing explainability mechanisms for lawyers. Developing fitting explanations can help lawyers, judges, and experts comprehend how an autonomous system came to a specific decision or action, making it easier to assess liability and determine causation. We recapitulate our identified benefits of and paths towards developing system explanations for lawyers.

Allocating responsibility. Explanations can help establish a causal link between an autonomous system's actions and the resulting harm, aiding in tort claims against producers and users. There are inherent limitations to providing explanations for complex autonomous systems. In some cases, it may be challenging to offer intuitive explanations, and it may not always be necessary for the court to have a detailed understanding of what exactly transpired in order to attribute liability. The key question becomes whether an end-user is responsible for the specific situation, if the producer marketed the product responsibly, if the system was adequately tested for the environment, and if its response was unpredictable.

Promoting accountability. It is important to avoid allowing system producers to use the defence of uncertainty by claiming, "It's Artificial Intelligence, so we don't know". Producers should be required to explain how they addressed and mitigated potential unpredictability. By promoting transparency and accountability, we can ensure that the responsibility lies with those involved in developing and deploying autonomous systems, rather than allowing ambiguity to be used as a defence in legal proceedings.

Adapting legal requirements. Next to explanations of opaque and complex systems, courts can be aided by adapting legal requirements, such as negligence or product defect standards. Some potential adjustments include expanding the duty of care owed by end-users to include a responsibility to oversee the autonomous system [38]. Explainability is also a central topic in the EU AI Act, which is currently in the process of being finalised.[3]

Explainability as a standard. In product liability cases, the concept of defect may need to be revised to consider the feasibility of alternative system designs with higher levels of explainability. [26] If such alternatives exist and are deemed reasonable, it could strengthen the argument for product defect claims. Legal standards for proving causation in the context of system-related harm could be adjusted to consider the unique characteristics of these systems. Courts could recognise the importance of system inspection and explanation in establishing a causal link between system behaviour and the harm suffered. If the issues in distinguishing causation and correlation could be resolved then counterfactual explanation techniques might be of significant value here. For introducing explainability as a standard for system design, there is also a close link to an ethical viewpoint on explainability requirements as it is sketched in [39]. A template for engineering explanations that fulfil ethical requirements is provided in their follow-up [40]. In future work, this template could be enhanced to also include the legal requirements that we identified in this vision.

To summarise, explainability and changes to legal requirements of autonomous systems can help mitigate the challenges associated with opaque autonomous systems in civil lawsuits. Improving explainability requirements and revising legal standards must be done with careful consideration of the interplay between them. By promoting transparency, accountability, and the ability to establish causation, these measures can contribute to a fairer and more effective legal system in the context of litigation cases for autonomous systems.

[6]Proposal for a Directive Of The European Parliament And Of The Council on liability for defective products, COM(2022)

References

[1] L. Chazette, W. Brunotte, and T. Speith, "Explainable software systems: from requirements analysis to system evaluation," *Requir. Eng.*, vol. 27, no. 4, pp. 457–487, 2022. [Online]. Available: https://doi.org/10.1007/s00766-022-00393-5

[2] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0004370218305988

[3] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box: Automated decisions and the gdpr," *Harvard journal of law & technology*, vol. 31, pp. 841–887, 04 2018.

[4] M. C. BUITEN, "Towards intelligent regulation of artificial intelligence," *European Journal of Risk Regulation*, vol. 10, no. 1, p. 41–59, 2019.

[5] P. Hacker and J.-H. Passoth, "Varieties of ai explanations under the law. from the gdpr to the aia, and beyond," in *xxAI-Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers*. Springer, 2022, pp. 343–373.

[6] M. Winikoff, G. Sidorenko, V. Dignum, and F. Dignum, "Why bad coffee? explaining bdi agent behaviour with valuings," *Artificial Intelligence*, vol. 300, p. 103554, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0004370221001053

[7] F. Ziesche, V. Klös, and S. Glesner, "Anomaly detection and classification to enable self-explainability of autonomous systems," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2021, pp. 1304–1309.

[8] M. Blumreiter, J. Greenyer, F. J. C. Garcia, V. Klös, M. Schwammberger, C. Sommer, A. Vogelsang, and A. Wortmann, "Towards self-explainable cyber-physical systems," in *22nd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion*, 2019, pp. 543–548.

[9] H. M. Taylor, C. Jay, B. Lennox, A. Cangelosi, and L. Dennis, "Should AI systems in nuclear facilities explain decisions the way humans do? an interview study," in *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2022, pp. 956–962.

[10] N. Du, J. Haspiel, Q. Zhang, D. Tilbury, A. K. Pradhan, X. J. Yang, and L. P. Robert, "Look who's talking now: Implications of av's explanations on driver's trust, av preference, anxiety and mental workload," *Transportation Research Part C: Emerging Technologies*, vol. 104, pp. 428–442, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X18313640

[11] P. A. M. Ruijten, J. M. B. Terken, and S. Chandramouli, "Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior," *Multimodal Technol. Interact.*, vol. 2, no. 4, p. 62, 2018.

[12] L. A. Dennis and N. Oren, "Explaining BDI agent behaviour through dialogue," *JAAMAS*, 2022.

[13] H. Hastie, H. Cuayáhuitl, N. Dethlefs, S. Keizer, and X. Liu, *Extrinsic Versus Intrinsic Evaluation of Natural Language Generation for Spoken Dialogue Systems and Social Robotics*. Singapore: Springer Singapore, 2017, pp. 303–311. [Online]. Available: https://doi.org/10.1007/978-981-10-2585-3_24

[14] N. Nobani, F. Mercorio, and M. Mezzanzanica, "Towards an explainer-agnostic conversational xai," in *IJCAI*, 2021.

[15] A. Bairy, W. Hagemann, A. Rakow, and M. Schwammberger, "Towards formal concepts for explanation timing and justifications," in *30th IEEE International Requirements Engineering Conference Workshops, RE 2022 - Workshops, Melbourne, Australia, August 15-19, 2022*. IEEE, 2022, pp. 98–102. [Online]. Available: https://doi.org/10.1109/REW56159.2022.00025

[16] Y. Xu, J. Collenette, L. Dennis, and C. Dixon, "Dialogue explanations for rules-based ai systems," in *Proceedings of the 5th International Workshop on Explainable and Transparent AI and Multi-Agent Systems (EXTRAAMAS 2023)*, 2023.

[17] I. V. T. Society, "IEEE standard for transparency of autonomous systems," *IEEE Std 7001-2021*, pp. 1–54, 2022.

[18] A. F. T. Winfield, S. Booth, L. A. Dennis, T. Egawa, H. Hastie, N. Jacobs, R. I. Muttram, J. I. Olszewska, F. Rajabiyazdi, A. Theodorou, M. A. Underwood, R. H. Wortham, and E. Watson, "IEEE p7001: A proposed standard on transparency," *Frontiers in Robotics and AI*, vol. 8, p. 225, 2021. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2021.665729

[19] M. Schwammberger and V. Klös, "From specification models to explanation models: An extraction and refinement process for timed automata," in *Proceedings Fourth International Workshop on Formal Methods for Autonomous Systems (FMAS) and Fourth International Workshop on Automated and verifiable Software sYstem DEvelopment (ASYDE), FMAS/ASYDE@SEFM 2022, and Fourth International Workshop on Automated and verifiable Software sYstem DEvelopment (ASYDE)Berlin, Germany, 26th and 27th of September 2022*, ser. EPTCS, M. Luckcuck and M. Farrell, Eds., vol. 371, 2022, pp. 20–37. [Online]. Available: https://doi.org/10.4204/EPTCS.371.2

[20] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," in *KDD*, 2016.

[21] Y.-L. Chou, C. Moreira, P. Bruza, C. Ouyang, and J. Jorge, "Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications," *Information Fusion*, vol. 81, pp. 59–83, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253521002281

[22] M. Thayaparan, M. Valentino, D. Ferreira, J. Rozanova, , and A. Freitas, "∂-explainer: Differentiable convex optimization for explainable multi-hop natural language inference," *Trans. ACL*, 2022.

[23] A. Deeks, "The judicial demand for explainable artificial intelligence," *Columbia Law Review*, vol. 119, no. 7, pp. 1829–1850, 2019.

[24] M. A. Lemley and B. Casey, "Remedies for robots," *The University of Chicago Law Review*, vol. 86, no. 5, pp. 1311–1396, 2019.

[25] M. Buiten, A. de Streel, and M. Peitz, "The law and economics of ai liability," *Computer Law & Security Review*, vol. 48, p. 105794, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0267364923000055

[26] M. C. Buiten, "Product liability for defective ai," July 2023, sSRN ID 4515202.

[27] S. J. Schweikart, "Who will be liable for medical malpractice in the future? how the use of artificial intelligence in medicine will shape medical tort law," *Minn. JL Sci. & Tech.*, vol. 22, p. 1, 2020.

[28] M. Ananny and K. Crawford, "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability," *new media & society*, vol. 20, no. 3, pp. 973–989, 2018.

[29] Ł. Górski and S. Ramakrishna, "Explainable artificial intelligence, lawyer's perspective," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, 2021, pp. 60–68.

[30] M. C. Buiten, "Chancen und grenzen erklärbarer algorithmen im rahmen von haftungsprozessen," in *Regulierung für Algorithmen und Künstliche Intelligenz*. Nomos Verlagsgesellschaft mbH & Co. KG, 2021, pp. 149–174.

[31] W. D. Smart, C. M. Grimm, and W. Hartzog, "An education theory of fault for autonomous systems," *Notre Dame J. on Emerging Tech.*, vol. 2, p. 33, 2021.

[32] A. D. Selbst, "Negligence and ai's human users," *BUL Rev.*, vol. 100, p. 1315, 2020.

[33] P. H. Padovan, C. M. Martins, and C. Reed, "Black is the new orange: how to determine ai liability," *Artificial Intelligence and Law*, vol. 31, no. 1, pp. 133–167, 2023.

[34] H. Fraser, R. Simcock, and A. J. Snoswell, "Ai opacity and explainability in tort litigation," in *2022 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 185–196. [Online]. Available: https://doi.org/10.1145/3531146.3533084

[35] D. C. Vladeck, "Machines without principals: liability rules and artificial intelligence," *Wash. L. Rev.*, vol. 89, p. 117, 2014.

[36] Y. Benhamou and J. Ferland, "Artificial intelligence & damages: Assessing liability and calculating the damages," *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law, Forthcoming*, 2020.

[37] O. Rachum-Twaig, "Whose robot is it anyway?: Liability for artificial-intelligence-based robots," *U. Ill. L. Rev.*, p. 1141, 2020.

[38] R. Zuroff, "Recognizing operators' duties to properly select and supervise ai agents–a (better?) tool for algorithmic accountability," *Canadian Journal of Law and Technology*, vol. 19, no. 1, p. 93, 2023.

[39] N. Balasubramaniam, M. Kauppinen, K. Hiekkanen, and S. Kujala, "Transparency and explainability of ai systems: Ethical guidelines in practice," in *Requirements Engineering: Foundation for Software Quality: 28th International Working Conference, REFSQ 2022, Birmingham, UK, March 21–24, 2022, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2022, p. 3–18. [Online]. Available: https://doi.org/10.1007/978-3-030-98464-9_1

[40] N. Balasubramaniam, M. Kauppinen, A. Rannisto, K. Hiekkanen, and S. Kujala, "Transparency and explainability of ai systems: From ethical guidelines to requirements," *Information and Software Technology*, vol. 159, p. 107197, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584923000514