ALGORITHMIC TORTS: THE ROLE OF EXPLAINABILITY

José Luiz de Moura Faleiros Júnior [*]

Abstract: The purpose of this paper is to elucidate the intersection of algorithmic explainability and civil liability, exploring the implications of complex algorithms on legal responsibility. Algorithms, while not inherently intelligent, depend on data to present probabilistic predictions, differing significantly from human intuition. The core of this study lies in examining the limits of liability for damages caused by sophisticated algorithms, considering their inherent unpredictability. By analyzing the accountability framework proposed by scholars like Frank Pasquale, Mark Coeckelbergh, and Giovanni Comandé, this paper highlights the importance of data-informed duties and relational justifications as elements of the principle of explainability. It underscores the need for a proactive approach to risk management and the establishment of ethical standards for trustworthy AI. The discussion includes the necessity of regulatory guidelines that balance technological innovation with preventive measures that ensures transparency, predictability, and accountability in the deployment of algorithmic systems.

Keywords: algorithmic torts; explainability; civil liability; accountability; risk assessment.

INTRODUCTION

The consolidation of the so-called 'information society' marks a significant transition period, necessitating a restructuring of certain theoretical bases within the Science of Law. Algorithms, which form the backbone of this digital transformation, are not inherently intelligent. They rely on past data to make probabilistic predictions, a process fundamentally different from the intuitive, imaginative, and creative capacities of humans. This divergence raises critical questions within legal frameworks, particularly concerning civil liability.

Civil liability traditionally hinges on the imputation of a duty to make reparations for harm caused by one's actions. However, the advent of complex algorithms introduces new dimensions to this field. Despite being parameterized and developed according to the latest technological standards, these algorithms can still cause damage due to factors that may be partially or entirely unpredictable. This unpredictability challenges the conventional boundaries of liability and necessitates a reassessment of legal

[*] Ph.D, University of São Paulo, Brazil. Email: josefaleirosjr@outlook.com / ORCID iD: https://orcid.org/0000-0002-0192-2336

responsibilities in the context of algorithmic operations.

This paper aims to explore the intersection of algorithmic explainability and civil liability. The core objective is to understand how the principle of explainability can inform and restructure the framework of civil liability, particularly in relation to complex algorithms. By analyzing the theoretical constructions of notable scholars such as Frank Pasquale, Mark Coeckelbergh, and Giovanni Comandé, the paper will highlight the importance of accountability and proactive risk management in the development and deployment of algorithms.

This study employs a multi-faceted methodological approach to explore the intersection of algorithmic explainability and civil liability. The primary methods include a comprehensive literature review, theoretical analysis, and comparative study. By integrating these elements, the paper aims to contribute to the ongoing discourse on the legal implications of algorithmic decision-making.

I. UNDERSTANDING ALGORITHMS AND THEIR LIMITATIONS

While there is no doubt that the technological advancements of the 20th century have led to undeniable gains in efficiency and triggered a profound paradigm shift, it cannot be denied that incessant data flows generate concerns about the risks of hyperconnectivity[1].

Klaus Schwab describes several technological innovations with exciting disruptive potential: (i) implantable technologies; (ii) digital presence; (iii) vision as a new interface; (iv) wearable technologies; (v) ubiquitous computing; (vi) pocket-sized supercomputers; (vii) storage for all; (viii) the Internet of Things and for things; (ix) connected homes; (x) smart cities; (xi) big data and decision-making; (xii) self-driving cars; (xiii) artificial intelligence applied to decision-making; (xiv) artificial intelligence applied to administrative functions; (xv) the relationship between robotics and services; (xvi) the rise of cryptocurrencies; (xvii) the sharing economy; (xviii) the relationship between governments and blockchain; (xix) 3D printing and manufacturing; (xx) 3D printing and human health; (xxi) 3D printing and consumer products; (xxii) designed beings; (xxiii) neurotechnologies[2].

In all these examples, one can identify exciting aspects inherently related to the disruptive potential of these new technologies. However, the risks of their unregulated and excessive adoption are equally visible. This discussion brings civil liability back to the center of investigation, as the doctrinal structure of the protection of artificial intelligence algorithms is directly

[1] Greengard, Samuel. *The Internet of Things*. Cambridge: The MIT Press, 2015, 58.

[2] Schwab, Klaus. *A quarta revolução industrial*. Translated by Daniel Moreira Miranda. São Paulo: Edipro, 2016, 10.

related to the implementation of these increasingly automated and data-driven disruptive technologies.

The debate, in this context, is guided by the appropriate legal framework for liability regimes and the functions of civil liability applicable to each situation. When it comes to seeking innovation, consumer relations immediately come to mind, as "products liability laws must still function to protect the consumer from harm by encouraging businesses to act appropriately to mitigate against foreseeable risks."[3]

It is precisely the spectrum of damage foreseeability that intensifies or amplifies the risks involved in the entire development process of algorithm-based applications (implicating the manufacturer/producer) and also the use of these technologies (implicating the owner/user) [4]. Therefore, it is necessary to formulate strong responses to the supposed 'grey area of imputability' identified in these cases. Otherwise, as Ugo Pagallo warns, all parties involved in the production and use of these systems would assume the risks of civil liability for damages caused by these machines "24 hours a day."[5]

At the core of this issue lies a contrast between the concepts of risk and danger, both relevant to the preventive and precautionary functions of civil liability. Despite the apparent linguistic similarity of the terms, Mafalda Miranda Barbosa warns that legally, danger and risk should not be confused. "more than the verification of mere danger, considerations related to the idea that it is fair to hold accountable those who derive benefit from an activity that is likely to cause harm to third parties are often at stake."[6]

In summary, the more abstract concept of danger is insufficient to eliminate the need to demonstrate the subjective element (fault) in liability. It is essential to ascertain the causal link based on a broader understanding of risk. This leads to the notion of 'foreseeability,' which aligns better—with the current state of the art—with the preventive function.

---

[3] Swanson, Greg. "Non-autonomous Artificial Intelligence Programs and Products Liability: How New AI Products Challenge Existing Liability Models and Pose New Financial Burdens." *Seattle University Law Review* 42 (2019): 1201-1222, 1222.

[4] Antunes, Henrique Sousa. "Inteligência Artificial e Responsabilidade Civil: Enquadramento." *Revista de Direito da Responsabilidade*, Coimbra, 1 (2019): 139-154, 141-142

[5] "Therefore, under strict liability rules for vicarious responsibility, owners and users of robots would be held strictly responsible for the behaviour of their machines 24-h a day, whereas, at times, negligence-based liability would add up to (but never avert) such strict liability regime". Pagallo, Ugo. *The Laws of Robots: Crimes, Contracts, and Torts*. Law, Governance and Technology Series, vol. 10. Cham/Heidelberg: Springer, 2013, 132.

[6] Barbosa, Mafalda Miranda. *Liberdade vs. Responsabilidade: A Precaução como Fundamento da Imputação Delitual?* Coimbra: Almedina, 2006, 352, freely translated. Original excerpt: "mais do que a verificação do simples perigo, estão em causa amiúde considerações ligadas à ideia de que é justo responsabilizar aquele que retira um proveito de uma atividade que com toda a probabilidade poderá causar prejuízos a terceiros".

When analyzing algorithms applied to economic activities, doctrine already debates the nebulous contingencies that demand more specific regulation. The major challenge extends beyond regulation aimed at data protection, though this is an important first step, as it brings attention to the desired accountability[7].

Undoubtedly, the processing of large data sets reveals immense potential for technological exploration but also denotes significant risks. It should be noted that technological singularity in algorithms can be achieved—this will be further analyzed later—but, in summary, we have not yet reached the point where we can assert the existence of a true symbiosis between the biological and the technological, to the extent of having "intelligent machines" comparable to the complexity of the human mind. Certainly, algorithms still operate in the heuristic domain, although this debate cannot be ignored at the current stage of technological development[8].

These 'non-intelligent' structures (artificial unintelligences, as Meredith Broussard calls them[9]) are still incapable of perceiving and assimilating the world in all its complexity, with sensory perceptions, moral discernment, critical analysis of reality, and various other characteristics identified only in human beings. Algorithms, even when powered by machine learning, are still fallible and extremely prone to errors in representation and assimilation, which, due to the inherently mathematical nature of data processing, only highlight the challenge of reconciling civil liability and its classical institutes with this new reality, even prospectively. This underscores the importance of human-machine interaction, which should not cause undue concern, as Daniel and Richard Susskind emphasize, "(...) the most efficient future lies with machines and human beings working together. Human beings will always have value to add as collaborators with machines".[10]

---

[7] "Accountability is a concept with many dimensions. It has been characterized by scholars as being an "elusive" and even "chameleon-like" concept, because it can mean very different things to different people. In its core meaning, accountability refers to the existence of a relationship whereby one entity has the ability to call upon another entity and demand an explanation and/or justification for its conduct. Over time, different data protection instruments have advanced different types of accountability mechanisms. In the GDPR, the principle of accountability is mainly used to signal that controllers are not only responsible for implementing appropriate measures to comply with the GDPR, but must also be able to demonstrate compliance at the request of supervisory authorities". Van Alsenoy, Brendan. *Data Protection Law in the EU: Roles, Responsibilities and Liability*. Cambridge: Intersentia, 2019, 318.

[8] Henderson, Harry. *Artificial Intelligence: Mirrors for the Mind*. New York: Chelsea House, 2007, 152.

[9] Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge: The MIT Press, 2018, 7-8.

[10] Susskind, Richard, and Daniel Susskind. *The Future of Professions: How Technology Will Transform the Work of Human Experts*. Oxford: Oxford University Press, 2015, 293.

Ryan Abbott, in a work specifically dedicated to the topic, suggests treating humans and robots equally, making no distinction even for legal protection purposes. In a consequentialist approach, he argues that if a robot can perform a job previously relegated only to humans, it should receive the same treatment as a human would in case of failure. His analysis is explicitly centered on the postulate of efficiency and the goal of preventing market distortions, which clearly does not align with the current stage of technological development that has not yet achieved technological singularity[11].

As Caitlin Mulholland states, "for a person to be obliged to repair unjust damage, it is fundamental that they have the autonomy to act, i.e., the ability to recognize the legality or illegality of their conduct and, at the same time, the habitual ability to identify and foresee the potential harm it may cause."[12]

Despite transitional ideas and proposals, what matters now is recognizing that until the mentioned technological singularity is achieved, unusual models may provisionally protect certain legal situations weakened by a lack of clarity regarding their protection. This is the case with existential legal situations, such as those involving algorithmic biases.

II. LEGAL FRAMEWORK AND CIVIL LIABILITY

There is no doubt that the period of transition in which the apogee of the so-called 'information society' is consolidated reveals nuances that, for the Science of Law, require the restructuring of certain theoretical bases. That said, we know that algorithms are not "intelligent." On the contrary, in order to function, they still depend on the past to make probabilistic heuristic predictions (because, in short, what they do is process data) and, for this simple and absolutely obvious reason, they are very far removed from the intuitive, imaginative, and creative way in which human beings act.

There are legal institutes permeated by interesting discussions that arise from this. This is the case of civil liability, whose functions have been

[11] "Policymakers should be concerned with the functionality of machines and consequentialist benefits – what will result in the greatest social benefit from these technologies – in deciding how to legally treat AI. At the end of the day, people do not concern themselves with whether a self-driving Tesla with an unpredictable neural network, a self-driving Uber using Good Old-Fashioned AI, or a human driver is behind the wheel of a car coming toward them. They – we – simply do not want to be run over." Abbott, Ryan. *The Reasonable Robot: Artificial Intelligence and the Law*. Cambridge: Cambridge University Press, 2020, 141-142.

[12] Mulholland, Caitlin. "Responsabilidade Civil e Processos Decisórios Autônomos em Sistemas de Inteligência Artificial (IA): Autonomia, Imputabilidade e Responsabilidade." In *Inteligência Artificial e Direito: Ética, Regulação e Responsabilidade*, edited by Ana Frazão and Caitlin Mulholland, 331-345. São Paulo: Thomson Reuters Brasil, 2019, 332.

investigated for some time by specialized doctrine. Certainly, it is in this field of study that the great concern about the limits of imputation of the duty to make reparations arises from any failure caused by complex algorithms that, although parameterized and adequately developed in the current state of the art, can be vectors of damage caused by factors that may be totally or partially predictable.

We work with accountability in the context of algorithmic development (and its underlying risks), signaling an undeniable propensity to establish duties that restructure the specific framework of classic dogma and reinforce proactive or investigative work ex ante—which prevents damage—as a vector for the exploration of activities whose risk is identified in the very complexity of the algorithms. To this end, when analyzing the issue, I have referred[13] to the proposals of three important researchers, namely: (i) Frank Pasquale's theoretical constructions around the principle of explainability and 'data-informed duties'; (ii) Mark Coeckelbergh's expanded concept of explainability, which opens up space for the proposal of 'relational justifications'; (iii) Giovanni Comandé's multi-layered accountability, based on ethical parameters legally introduced into the regulatory debate.

We speak of "artificial intelligence" by convention, although the expression itself is confusing, since it has come to condense several meanings more akin to a multidisciplinary[14] branch of science than a technology per se. For this reason, the parameterization of more accurate (and consequently more contingent) duties leads to the notion of 'predictability,' which better aligns—in the current state of the art—with the preventive function of civil liability.

Foreign doctrine uses the term foreseeability[15] to summarize this element even in contexts in which the theory of fault may make more sense (such as

---

[13] Faleiros Júnior, José Luiz de Moura. "Responsabilidade por falhas de algoritmos de inteligência artificial: ainda distantes da singularidade tecnológica, precisamos de marcos regulatórios para o tema?" *Revista de Direito da Responsabilidade*, Coimbra, vol. 4, 2022, 906-933.

[14] In 2020, with the publication of the 4th edition of their seminal work, Stuart J. Russell and Peter Norvig provided a more detailed account of this conceptual problem. See Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed. London: Pearson, 2020, 19-20. For an even more comprehensive historical perspective, it is essential to read the recent and absolutely elucidative work by Wooldridge, Michael. *The Road to Conscious Machines: The Story of AI*. Los Angeles: Pelican, 2020.

[15] "Foreseeability remains a necessary ingredient even where liability is otherwise "strict" (i.e., where no showing of negligence by the plaintiff is necessary to recovery). There will be situations, particularly as emergent systems interact with one another, wherein otherwise useful technology will legitimately surprise all involved. Should these systems prove deeply useful to society, as many envision, some other formulation than foreseeability may be necessary to assess liability". Calo, Ryan. "Robotics and the Lessons of Cyberlaw." *California Law Review* 103 (2015): 513-563, 555.

when investigating the negligent behavior of the developer of an algorithmic system). However, it is also recognized that it is necessary to go further in the search for an adequate criterion to meet the precautionary function of civil liability in contexts of total unpredictability.

Complex algorithms are expected to be designed from collaborative structures in which trust permeates the relationships between professionals and corporations directly involved in all stages of their development, signaling the importance of trust for AI (trustworthy AI), anchored in ethics, and which can be studied in the light of Giddens' sense of reliability when dealing specifically with abstract principles of technical knowledge[16].

The essence of the three postulates coined by Jack Balkin[17] shows that the expected degree of exceptional diligence on the part of the agent who programs/develops an algorithm stems not only from the expectation of compliance with legislation and risk management (compliance) but also from proactive action to mitigate risks (*ex ante* accountability), in keeping with the famous 'responsibility principle' advocated by Hans Jonas[18]. However, this does not rule out the importance of formulating rules of good practice and governance aimed at possible damage and dealing with it (*ex post* accountability), taking into account the nature, scope, purpose, probability, and severity of the risks and benefits arising from the adoption of algorithmically structured systems.

The cooperative nature identified in Balkin's proposals led Frank Pasquale to propose a 'fourth law': a robot must always indicate the identity of its creator, controller, or owner[19]. This new postulate concretizes what the

---

[16] According to Giddens, "(…) confidence in the reliability of a person or system, regarding a given set of outcomes or events, where that confidence expresses faith in the probity or love of another, or in the correctness of abstract principles (technical knowledge)". Giddens, Anthony. *The Consequences of Modernity*. Stanford: Stanford University Press, 1990, 34.

[17] The centrality of this new concern must be—as it always has been—the human aspect. This conclusion is drawn, for example, from reading the classic Three Laws of Robotics described by Isaac Asimov in the short story "Runaround" from his famous collection *I, Robot* (Asimov, Isaac. *Eu, Robô*. Translated by Aline Storto Pereira. São Paulo: Aleph, 2014), which in turn inspired the American Jack Balkin to also formulate three postulates for the legal framework of this complex discussion. These have been referred to by the doctrine as the 'laws of robotics in the era of Big Data' and are summarized as follows: (a) algorithmic operators must be fiduciaries of information regarding their clients and end-users; (b) algorithmic operators have duties towards the general public; (c) algorithmic operators have a public duty not to engage in algorithmic nuisances. Balkin, Jack M. "The Three Laws of Robotics in the Age of Big Data." *Ohio State Law Journal* 78 (August 2017): 1-45. Accessed February 2, 2024. http://ssrn.com/abstract=2890965.

[18] Jonas, Hans. *Le Principe Responsabilité: Une Éthique pour la Civilisation Technologique*. Translated by Jean Greisch. 3rd ed. Paris: Éditions du Cerf, 1992, 225.

[19] Pasquale, Frank. "Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society." *University of Maryland Legal*

doctrine already supported as the 'principle of explainability,' aimed at allowing a given machine to indicate who its creator is and, eventually, also reveal the identity of its owner or user/operator.

If risk is the central element of all these proposals, the notion of danger—and precaution itself—also instigates reflections on the repercussions of unbridled technological development. It must be stressed that the concept, typology, and severity of damage that inspire the formatting of civil liability systems over time have varied from a perspective proportional to the very transformation of society[20].

Undeniably, the fact that the risk is proven or potential does not rule out the relevance of the principles of prevention and precaution, precisely because any kind of "new damage" generates a certain excitement and, as Ulrich Beck warns, leads to assumptions of social acceptance of new technologies[21]—even if not fully tested—due to the fact that risk, to some extent, becomes inherent in the various activities of everyday life.

Pasquale's proposal also finds support in the classic work of Stuart Russell and Peter Norvig, who already spoke of the 'quantification of uncertainty': "Agents in the real world may need to handle uncertainty, whether due to partial observability, nondeterminism, or adversaries. An agent may never know for sure what state it is in now or where it will end up after a sequence of actions."[22] In summary, it can be said that the conjectures from which the 'data-informed duties' are conceived are in line with the aforementioned 'fourth law of robotics' proposed by Pasquale (principle of explainability)[23], since his idea reinforces the need to overcome a problem also described by the author in another work[24]: that of algorithmic 'black boxes,' usually identified by the use of machine learning techniques that are opaque or non-transparent to users[25].

In civil liability, a very close connection with the principle of

---

Studies Research Papers 21 (2017): 1-13.

[20] Venturi, Thaís G. Pascoaloto. *Responsabilidade Civil Preventiva: A Proteção Contra a Violação dos Direitos e a Tutela Inibitória Material*. São Paulo: Malheiros, 2014, 248.

[21] Beck, Ulrich. *Risk Society: Towards a New Modernity*. Translated by Mark Ritter. London: Sage Publications, 1992. 6.

[22] Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed. London: Pearson, 2020, 403.

[23] Pasquale, Frank. "Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society." *University of Maryland Legal Studies Research Papers* 21 (2017): 1-13.

[24] Pasquale, Frank. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge: Harvard University Press, 2015, 6-7.

[25] Asaro, Peter M. "A Body to Kick, But Still No Soul to Damn: Legal Perspectives on Robotics." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney, and George A. Bekey, 169-186. Cambridge: The MIT Press, 2011, 169-186.

explainability can be traced in a number of proposals. One of them, by Mark Coeckelbergh, relies on 'relational justifications' to broaden the notion of explainability to the point where experts are required to develop algorithms for AI systems clearly, indicating that they are capable of developing them, that they wish to do so, and explaining the reasons that led them to make each decision along the way, including for training and testing stages of data sets[26].

An alternative proposal to the structuring of data-informed duties involves the recognition of the preventive function and its functionalization based on the desirable accountability[27] for the development of algorithms. This idea is taken from the writings of Giovanni Comandé, who stresses the need for a transition from traditional strict liability to a model of accountability that also deals with preventive and precautionary functions, imposing on those who assume a better hierarchical position in terms of taking on "informed" duties the duty to make choices and justify them to those on whom their effects fall.

III. EXPLAINABILITY'S ROLE FOR TRUSTWORTHY AI

In the information society, accountability must represent a 'culture'[28] (rather than a 'duty') and its effects must go beyond mere 'accountability' (as one might think from the literal translation of the term) for the choices made. It is, therefore, the legitimate expectation that the agent will answer, in the appropriate spheres (political, civil, criminal, administrative, ethical, social), for their possible failures and the shortcomings of their choices (especially if they are informed choices)[29].

With the protection of trust as a priority, the discussion on accountability includes ethical aspects that are now being reinterpreted for the application

---

[26] Coeckelbergh, Mark. "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability." *Science and Engineering Ethics* 26 (2020): 2051-2068, 2066.

[27] The term 'responsibility' does not have a single meaning. Its scope is even broader in languages like French or Spanish, where 'responsibility' is used in relation to a wide range of legal, political, and economic relations and within them, their respective dimensions. In English, however, the existence of different terms to refer to the various dimensions of responsibility—responsibility, accountability, liability—allows for a more precise application of the concept. For more on this, see Rosenvald, Nelson. "A Polissemia da Responsabilidade Civil na LGPD." *Migalhas de Proteção de Dados*, November 6, 2020. Accessed February 2, 2024. https://s.migalhas.com.br/S/477BB2.

[28] Nissenbaum, Helen. "Accountability in a Computerized Society." *Science and Engineering Ethics* 2, no. 1 (March 1996): 5-42, 7.

[29] Comandé, Giovanni. "Intelligenza Artificiale e Responsabilità tra Liability e Accountability: Il Carattere Trasformativo dell'IA e il Problema della Responsabilità." In *Analisi Giuridica dell'Economia. Studi e Discussioni sul Diritto dell'Impresa*, edited by Antonio Nuzzo and Gustavo Olivieri, 185-200. Bologna: Il Mulino, 2019, 185.

of artificial intelligence systems in the field of research that concerns the need for specific regulation. Instead of regulating "how" algorithms should be created, applied, and monitored, this more open model creates specific niches in which it makes more sense to establish more open guidelines, with a degree of generalization sufficient to guide technological development without depriving it of the environment conducive to its free testing and pivoting.

In short, this is a summary of the considerations needed to understand the appropriate legal framework for explainability in relation to civil liability:

a) A first observation on the subject is that, with regard to the requirements imposed on algorithmic operators[30], duties of care are imposed, which concern: a.1) the choice of technology, particularly in light of the tasks to be carried out and the operator's own skills and abilities; a.2) the organizational framework provided, especially with regard to adequate monitoring; and a.3) maintenance, including any safety checks and repairs. Failure to comply with these obligations could trigger liability for guilt (even if presumed, reversing the burden of proof), regardless of whether the operator is also strictly responsible for the risk created by implementing a certain technology, since they are not dealing with the uncertain or with high or excessive degrees of risk.

b) With regard to producers and manufacturers, in consumer relations, including those who act incidentally as operators, the following standards of conduct must be observed, according to Benhamou and Ferland[31]: b.1) design, describe, and market products in a way that allows them to comply with 'data-informed duties,' making risks more foreseeable (with an emphasis on foreseeability); and b.2) adequately monitor the product after it has been put into circulation, in light of the characteristics of emerging digital technologies, in particular their openness and dependence on the general digital environment, including obsolescence (programmed or not), the emergence of malware, or even their vulnerability to possible external attacks.

c) The so-called supervision, understood in the context of the duty to monitor ("superior" or "hierarchical," which can even be the result of state police power, in what Pasquale calls "oversight"[32]), could be achieved by

---

[30] The author is assertive: “We might hold many different potential actors liable, including the owner, operator, retailer, hardware designer, operating system designer, or programmer(s), to name only a few possibilities” Balkin, Jack M. "The Path of Robotics Law." *California Law Review Circuit* 6 (June 2015): 45-60, 52.

[31] Benhamou, Yaniv, and Justine Ferland. "Artificial Intelligence & Damages: Assessing Liability and Calculating the Damages." In *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law*, edited by Pina D’Agostino, Carole Piovesan, and Aviv Gaon, 196-197. Toronto: Thomson Reuters Canada, 2021, 196-197.

[32] Pasquale, Frank. *New Laws of Robotics: Defending Human Expertise in the Age of AI*. Cambridge: Harvard University Press, 2020, 99.

carrying out audits and studies of the specific algorithm, even after its release to the market. Thus, as a result of the implementation of supervised risk assessment systems, it would be expected that anomalies would be identified and that systems would be parameterized in advance to "warn" of the occurrence of unexpected behaviors, as well as observing specific trends of evolution based on machine learning to "predict" anomalous behaviors. Once such monitoring has been implemented, the obligation to inform potential victims arises as an annex to the duty of objective good faith[33], convoluted into the principle of transparency, which has a close correlation with the principle of explainability.

d) If feasible, it is argued that producers, manufacturers, or developers should be compelled to include mandatory backdoors[34] in their AI systems. Other designations for this are the expressions "emergency brakes by design," "shut down features," or features that allow operators or users to "turn off the AI" by manual commands or make it 'unintelligent' by pressing a "panic button."[35] Failure to guarantee such tools and options could be considered a design defect sufficient to justify liability for breach of the general duty of safety that would be imposed on them, which would open up the possibility of protecting damages through product liability, recognizing the algorithm itself as defective. In fact, depending on the circumstances, manufacturers or developers could also be obliged to "turn off" the AI systems themselves as part of their algorithmic monitoring and auditing tasks.

e) Similar to the after-sales duties already known in the consumer market and consisting of warnings and instructions to recall defective products, producers/manufacturers can also assume support and correction duties—corollaries of auditability and transparency[36]—in line with other recent developments on the potential obligation of software developers to update unsafe algorithms for as long as said technology is on the market (i.e., beyond any contractual stipulation on warranty period) [37].

---

[33] Wischmeyer, Thomas. "Artificial Intelligence and Transparency: Opening the Black Box." In *Regulating Artificial Intelligence*, edited by Thomas Wischmeyer and Timo Rademacher, 75-89. Cham: Springer, 2020, 76.

[34] Liao, Cong, Haoti Zhong, Anna Squicciarini, et al. "Backdoor Embedding in Convolutional Neural Network Models via Invisible Perturbation." *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy*, March 2020, 97-108. Accessed February 2, 2024. https://doi.org/10.1145/3374664.3375751, 99-105.

[35] Benhamou, Yaniv, and Justine Ferland. "Artificial Intelligence & Damages: Assessing Liability and Calculating the Damages." In *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law*, edited by Pina D'Agostino, Carole Piovesan, and Aviv Gaon, 196-197. Toronto: Thomson Reuters Canada, 2021, 196-197.

[36] Pasquale, Frank. "Data-Informed Duties in AI Development." *Columbia Law Review* 119 (2019): 1917-1940, 1937.

[37] In fact, although no law clearly contains an explicit obligation to do so, there is already jurisprudential signaling that interprets existing legal norms in a way that creates such an

CONCLUSION

Although the classic theories of civil liability applied to the medical field represent a coherent set of rules for the activity, dividing the debate between the characters patient, doctor, hospital and manufacturer, when we add a fifth character - artificial intelligence - the contours of this division of responsibility can change. In conclusion, the intersection of algorithmic explainability and civil liability presents a critical juncture in the evolving landscape of legal responsibility. As algorithms become increasingly integrated into the fabric of the information society, the traditional frameworks of civil liability must adapt to address the unique challenges posed by these complex systems. The inherent unpredictability of algorithms necessitates a reevaluation of liability principles to ensure that the legal system can effectively manage the risks associated with their deployment.

The principle of explainability emerges as a cornerstone in this adaptation, providing a basis for understanding and mitigating the potential harms caused by algorithmic decisions. Scholars such as Frank Pasquale, Mark Coeckelbergh, and Giovanni Comandé have emphasized the importance of accountability frameworks that include data-informed duties and relational justifications. These frameworks underscore the necessity for proactive risk management strategies and the establishment of ethical standards to foster trustworthy AI systems. By incorporating explainability into the legal framework, we can enhance transparency and predictability, thereby safeguarding civil rights and promoting public trust in technological innovations.

Moreover, the need for comprehensive regulatory guidelines is paramount. These guidelines should balance the promotion of technological innovation with preventive measures to protect against foreseeable risks. The discussion highlights the imperative for a legal structure that not only addresses the current capabilities and limitations of AI but also anticipates future developments. This proactive approach ensures that as technology evolves, the legal system remains robust and capable of protecting individuals and society from potential harms.

Ultimately, in this paper, I advocate for a dynamic and flexible legal framework that can adapt to the rapid advancements in technology while maintaining the core principles of accountability and responsibility. By embracing the principle of explainability and fostering a culture of

---

obligation for developers; this is the case with the Dutch precedent Consumentenbond v. Samsung. For a detailed study of the case and its repercussions, see: Wolters, Pieter T. J. "The Obligation to Update Insecure Software in the Light of Consumentenbond v. Samsung." *Computer Law & Security Review* 35, no. 3 (May 2019): 295-305.

transparency, we can navigate the complexities of algorithmic torts and ensure that the deployment of AI systems contributes positively to society. As we move forward, the integration of these principles into the legal system will be crucial in managing the balance between innovation and protection, ensuring that technological progress does not come at the expense of human rights and ethical standards.

REFERENCES

Abbott, Ryan. *The Reasonable Robot: Artificial Intelligence and the Law*. Cambridge: Cambridge University Press, 2020.

Antunes, Henrique Sousa. "Inteligência Artificial e Responsabilidade Civil: Enquadramento." *Revista de Direito da Responsabilidade*, Coimbra, 1 (2019): 139-154.

Asaro, Peter M. "A Body to Kick, But Still No Soul to Damn: Legal Perspectives on Robotics." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney, and George A. Bekey, 169-186. Cambridge: The MIT Press, 2011.

Asimov, Isaac. *Eu, Robô*. Translated by Aline Storto Pereira. São Paulo: Aleph, 2014.

Balkin, Jack M. "The Path of Robotics Law." *California Law Review Circuit* 6 (June 2015): 45-60.

Balkin, Jack M. "The Three Laws of Robotics in the Age of Big Data." *Ohio State Law Journal* 78 (August 2017): 1-45. Accessed February 2, 2024. http://ssrn.com/abstract=2890965.

Barbosa, Mafalda Miranda. *Liberdade vs. Responsabilidade: A Precaução como Fundamento da Imputação Delitual?* Coimbra: Almedina, 2006.

Beck, Ulrich. *Risk Society: Towards a New Modernity*. Translated by Mark Ritter. London: Sage Publications, 1992.

Benhamou, Yaniv, and Justine Ferland. "Artificial Intelligence & Damages: Assessing Liability and Calculating the Damages." In *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law*, edited by Pina D'Agostino, Carole Piovesan, and Aviv Gaon, 196-197. Toronto: Thomson Reuters Canada, 2021.

Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge: The MIT Press, 2018.

Calo, Ryan. "Robotics and the Lessons of Cyberlaw." *California Law Review* 103 (2015): 513-563.

Coeckelbergh, Mark. "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability." *Science and Engineering Ethics* 26 (2020): 2051-2068.

Comandé, Giovanni. "Intelligenza Artificiale e Responsabilità tra Liability e

Accountability: Il Carattere Trasformativo dell'IA e il Problema della Responsabilità." In *Analisi Giuridica dell'Economia. Studi e Discussioni sul Diritto dell'Impresa*, edited by Antonio Nuzzo and Gustavo Olivieri, 185-200. Bologna: Il Mulino, 2019.

Faleiros Júnior, José Luiz de Moura. "Responsabilidade por falhas de algoritmos de inteligência artificial: ainda distantes da singularidade tecnológica, precisamos de marcos regulatórios para o tema?" *Revista de Direito da Responsabilidade*, Coimbra, vol. 4, 2022, 906-933.

Giddens, Anthony. *The Consequences of Modernity*. Stanford: Stanford University Press, 1990.

Greengard, Samuel. *The Internet of Things*. Cambridge: The MIT Press, 2015.

Henderson, Harry. *Artificial Intelligence: Mirrors for the Mind*. New York: Chelsea House, 2007.

Jonas, Hans. *Le Principe Responsabilité: Une Éthique pour la Civilisation Technologique*. Translated by Jean Greisch. 3rd ed. Paris: Éditions du Cerf, 1992.

Liao, Cong, Haoti Zhong, Anna Squicciarini, et al. "Backdoor Embedding in Convolutional Neural Network Models via Invisible Perturbation." *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy*, March 2020, 97-108. Accessed February 2, 2024. https://doi.org/10.1145/3374664.3375751.

Mulholland, Caitlin. "Responsabilidade Civil e Processos Decisórios Autônomos em Sistemas de Inteligência Artificial (IA): Autonomia, Imputabilidade e Responsabilidade." In *Inteligência Artificial e Direito: Ética, Regulação e Responsabilidade*, edited by Ana Frazão and Caitlin Mulholland, 331-345. São Paulo: Thomson Reuters Brasil, 2019.

Nissenbaum, Helen. "Accountability in a Computerized Society." *Science and Engineering Ethics* 2, no. 1 (March 1996): 5-42.

Pagallo, Ugo. *The Laws of Robots: Crimes, Contracts, and Torts*. Law, Governance and Technology Series, vol. 10. Cham/Heidelberg: Springer, 2013.

Pasquale, Frank. "Data-Informed Duties in AI Development." *Columbia Law Review* 119 (2019): 1917-1940.

Pasquale, Frank. "Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society." *University of Maryland Legal Studies Research Papers* 21 (2017): 1-13. Accessed February 2, 2024. http://ssrn.com/abstract=3002546.

Pasquale, Frank. *New Laws of Robotics: Defending Human Expertise in the Age of AI*. Cambridge: Harvard University Press, 2020.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge: Harvard University Press, 2015.

Rosenvald, Nelson. "A Polissemia da Responsabilidade Civil na LGPD." *Migalhas de Proteção de Dados*, November 6, 2020. Accessed February 2, 2024. https://s.migalhas.com.br/S/477BB2.

Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed. London: Pearson, 2020.

Schwab, Klaus. *A quarta revolução industrial*. Translated by Daniel Moreira Miranda. São Paulo: Edipro, 2016.

Susskind, Richard, and Daniel Susskind. *The Future of Professions: How Technology Will Transform the Work of Human Experts*. Oxford: Oxford University Press, 2015.

Swanson, Greg. "Non-autonomous Artificial Intelligence Programs and Products Liability: How New AI Products Challenge Existing Liability Models and Pose New Financial Burdens." *Seattle University Law Review* 42 (2019): 1201-1222.

Van Alsenoy, Brendan. *Data Protection Law in the EU: Roles, Responsibilities and Liability*. Cambridge: Intersentia, 2019.

Venturi, Thaís G. Pascoaloto. *Responsabilidade Civil Preventiva: A Proteção Contra a Violação dos Direitos e a Tutela Inibitória Material*. São Paulo: Malheiros, 2014.

Wischmeyer, Thomas. "Artificial Intelligence and Transparency: Opening the Black Box." In *Regulating Artificial Intelligence*, edited by Thomas Wischmeyer and Timo Rademacher, 75-89. Cham: Springer, 2020.

Wolters, Pieter T. J. "The Obligation to Update Insecure Software in the Light of Consumentenbond v. Samsung." *Computer Law & Security Review* 35, no. 3 (May 2019): 295-305.

Wooldridge, Michael. *The Road to Conscious Machines: The Story of AI*. Los Angeles: Pelican, 2020.

* * *

José Luiz de Moura Faleiros Júnior
PhD in Civil Law from the University of São Paulo (USP/Largo de São Francisco). Currently pursuing a PhD in Law, specializing in 'Law, Technology, and Innovation' at the Federal University of Minas Gerais (UFMG). Holds a Master's and Bachelor's degree in Law from the Federal University of Uberlândia (UFU). Specialist in Digital Law. Lawyer. Professor.
Email: josefaleirosjr@outlook.com
ORCID iD: https://orcid.org/0000-0002-0192-2336

10.59224/bjlti.v2i1.210-224
ISSN: 2965-1549