Liability for AI Driven Systems

Ana Taveira da Fonseca, Elsa Vaz de Sequeira, and Luís Barreto Xavier

Abstract This article tries to assess if the current civil liability regimes provide a sound framework to tackle damages when AI systems—especially those based on machine-learning—are involved. We try to find answers for three questions: is there a place for fault-based liability, when it is impossible to ascertain, among multiple actors, whose action caused the damage? Are current strict liability regimes appropriate to address no-fault damages caused by the functioning of AI-systems or a new system is needed? When should an agent be exempted from liability? This analysis takes into consideration the important work produced within the European Union, especially the 2019 Report on "Liability for AI and Other Emerging Digital Technologies" (by the Expert Group set up by the European Commission), the European the Parliament 2020 Resolution on Civil Liability for AI, the 2021 Draft AI Act, the 2022 Draft AI Liability Directive and the 2022 Draft Product Liability Directive.

1 Presentation of the Problems

Digital technologies are evolving at a fast pace and artificial intelligence (AI) impacts all sectors of the economy and contemporary life. The operation of modern standalone or software based AI systems is likely to be associated to harm. In this article, we address the question of whether the traditional responses to the problem of compensation through civil liability are adequate to tackle damages when AI systems are put in place. While we depart from a Civil Law jurisdiction point of view, our discussion tries to go beyond the boundaries of our own legal tradition.

The challenges the traditional civil liability regimes face because of the dissemination of AI systems are linked to specific features of the operation of such

A. T. da Fonseca · E. Vaz de Sequeira (✉) · L. Barreto Xavier
Universidade Católica Portuguesa, Law School, Católica Research Centre for the Future of Law, Lisbon, Portugal
e-mail: anatf@ucp.pt; evs@ucp.pt; lbx@ucp.pt

systems: the ability of AI systems to making decisions in a growingly autonomous manner (Turner 2019, pp. 70–75; Ebers 2020, pp. 46–48; Chesterman 2021, pp. 31–62); the opacity of the machine learning based technologies (Ebers 2020, pp. 48–50; Chesterman 2021, pp. 63–82); the involvement of various agents in building, assembling, introducing into the market, customizing, selecting and supervising the data, training, updating and using the system (Expert Group on Liability and New Technologies—New Technologies Formation, Liability for artificial intelligence and other emerging technologies, Directorate-General for Justice and Consumers (European Commission) 2019, p. 35); the vulnerability to cyberattacks. All these factors contribute to the difficulty of deciding who—if anyone—should respond for a loss or harm. Hence, unless we refine or rethink traditional approaches, those who suffer damages are likely to be deprived of a fair compensation.

A case submitted under the traditional fault based liability against an operator or user of an AI system is very hard to succeed. The causal process is typically unknown to the victim. The black box effect of machine learning algorithms obstruct the transparency and explainability of the decision-making process. The number of agents potentially involved add to the complexity of the task. The plaintiff's burden of evidencing the fault is, most of the times, impossible to accomplish.

In this article, we consider three questions.

The first regards the possibility of establishing fault-based liability when various actors involved in the process disregarded the applicable rules, but it is impossible to determine which of the actions constitutes the actual cause of the damage.

Second question: when no fault has been committed and the damage is due to the functioning of AI-systems, should we apply any of the strict liability regimes in force? Should we, instead, design a specific regime for damages associated to AI systems?

Third and last question: when should the liability of the agent be excluded? In other words, what are the defenses the agent is able to put forward to escape liability?

The European Union (EU) has published important documents dealing with AI and civil liability on a general basis.

The Report on "Liability for AI and Other Emerging Digital Technologies" (2019 Report), presented by the Expert Group on Liability and New Technologies—New Technologies Formation (Expert Group), set up by the European Commission, discusses the application of existing liability regimes to emerging digital technologies, with a focus on AI, and the need for reform of those regimes.

The European Parliament (EP) has adopted on 20 October 2020 a "Resolution with Recommendations to the Commission on a Civil Liability Regime for AI" (2020 EP Resolution), including a Proposal for a Regulation of the European Parliament and the Council on Liability for the Operation of AI-Systems.

The 2021 Draft AI Act[1] does not address the civil liability issues. Instead, the European Commission has proposed in 2022 a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (the 'AI Liability Directive') [2] and, at the same time, a new Directive on Liability for Defective Products, replacing the old Product Liability Directive (PLD). [3]

The 2019 Report distinguishes two types of potential perpetrators. The frontend operator is the "natural or legal person who exercises a degree of control over a risk connected with the operation and functioning of the AI-system and benefits from its operation". The backend operator is the "natural or legal person who, on continuous basis, defines the features of the technology, provides data and essential backend support service and therefore also exercises a degree of control over the risk connected with the operation and functioning of the AI-System".[4] The grounds for the liability of these agents seem to lie on the position of control that they exercise over the AI system and, in relation to the frontend operator, also in the benefit that he obtains from it. Although it is not clear, the producer, the programmer and the person in charge of feeding the system would act as backend operators. It seems more difficult to identify who might take on the role of frontend operator, particularly as the report is ambiguous on this topic. If, on one hand, the users of these systems seem to be comprised here, on the other hand recital 11 of the proposed regulation states that the user should only (objectively) respond if he is an operator. This suggests that there are users who are operators—those who, in addition to benefitting from the use, exercise some control over the process—and others who are not, because they do not enjoy similar attributes.

It should be pointed out that the concepts of frontend and backend operator are used here in a manner that do not fully coincide with the European Parliament's proposal. Under this approach, the fronted operator is "the person primarily deciding on and benefitting from the use of the relevant technology" and the backend operator is "the person continuously defining the features of the relevant technology and providing essential and ongoing backend support".[5] It appears that the concept of fronted operator proposed in the Expert Group's study is broader and at the same time clearer than the one used in the European Parliament's proposal. Both require the user of the AI system to derive a benefit from its use. Yet, whereas the latter requires, in addition, control over the source of danger constituted by that system, the former is satisfied with the decision to use that source of danger.

---

[1] European Commission 21-4-2021 Proposal for a Regulation of the EP and the Council Laying Down Harmonised Rules on AI (AI Act) and Amending Certain Union Legislative Acts.

[2] 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), from 28-9-2022.

[3] Proposal for a Directive of the European Parliament and of the Council on Liability for Defective Products, from 28-9-2022.

[4] Article 3 e) and f).

[5] Point 11.

2 Subjective Liability in Case of Alternative Causation

The involvement of an AI system in the production of damages increases the difficulty of establishing fault and causation.

It will not always be easy to demonstrate the existence of a subjective imputation of the damage to the agent, due to lack of purpose or negligence—because the novelty of this type of situation has not yet allowed the development of duties of care-, just as it may be difficult to evaluate the agent's culpability. The autonomy of these systems makes it impossible, to a greater or lesser extent, to foresee how they will act in a specific case (Chesterman 2021, pp. 31–38, 60–62). The lack of predictability compromises the ability to make a prognosis as to the possible results of the conduct. This, in turn, may hinder the assessment of the culpability of the agent for having created or used the AI in those concrete circumstances (Barbosa 2020, p. 284).

In the same way, it would be extremely difficult to proceed with the objective imputation of the damage to the conduct of one of the participants in the process, due to the impossibility of ascertaining the actual cause of the damage. One of the possible outcomes at this level is the conclusion that any of the participations in the process of creation or use of the AI system could have produced the damage. In other words, any of them could be at the origin of the damage, but it is not known which of them is actually responsible for the damage.

The solution to the alternative causation problem is much discussed in legal theory, and it is debated whether one should:

(a) simply rule out the liability of the agents, for lack of causation;
(b) exclude in such cases the requirement of causation (Bydlinski 1959, pp. 6 et seq);
(c) replace *actual causation* with *possible causation*. Instead of demonstrating that the action of each agent was the actual cause of the damage, it would be sufficient to prove that such an action was a possible or potential cause of that damage. At the same time, it may be established a reversal of the burden of proof with regard to this assumption, presuming the existence of potential causation (Larenz 1994, pp. 571–572).[6]

The majority of authors lean towards the last option, rejecting the first one as, in balancing the interests of the potential injurer and the affected person, the former is privileged. It does not seem defendable if we think that any of the agents performed an action able of causing the damage and may have actually caused it. The second option is rejected as it establishes an unnecessary and not commendable deviation to the rule of a fault-liability system (Brambring 1973, p. 59; Larenz 1994, p. 571; Wagner 2018, p. 2318). When the requirement of causation is discarded, both the actual causality of the action and its suitability to produce the harm are not

---

[6] It makes no sense to presume the actual causality of such actions, since the reason for making use of the presumption is precisely the failure to determine the actual cause of the damage.

evaluated. That may lead to the liability of someone who did not practice an act capable of generating the damage. The third option, on the other hand, besides being based on the prevalence of the interest of the affected person over the conflicting interest of the potential injurer, does not involve the risk of holding liable someone who couldn't contribute to the event. The presumption of potential causation is a way of lightening the requirements of proof in this matter, protecting the affected person, since it is hard to prove the adequacy of each individual participation to the production of the entire result (Brambring 1973, pp. 95 et seq).

It should be emphasized that this lightening of the burden of proof of causality does not exempt the fulfilment of the other elements of civil liability for each agent involved in the causal process. Even in relation to causation, as has been said, the adequacy of the individual behavior to the production of the whole damage should be proved, unless the legislator has established a presumption of adequacy in order to protect the position of the affected person. In such a case, liability may be excluded if the potential perpetrator proves that his or her conduct did not cause the damage, that the conduct of another agent caused the damage, or that in the present case there was a ground of justification, exculpation or even impunity, which benefited him or one of the other agents involved (Larenz 1994, pp. 573 et seq, 576–578; Staudinger and Eberl-Borges 2018, pp. 29–30, 41–42; Wagner 2020, p. 2321 et seq).

It is sometimes questioned whether the application of this regime of joint and several liability of all the agents should be dependent on the verification of three requirements: the existence of a chronological connection between the individual conducts, the presence of a spatial connection between those same conducts and/or the common nature of the actions performed by each agent. In addition to this objective connection, it is also questionable whether a subjective connection should be imposed, i.e. the need that all agents be aware of each other or, in a less strict version, that they should be aware of each other (Brambring 1973, pp. 62 et seq; Larenz 1994, p. 574).

The absence of joint participation in these cases—characterized by a bilateral awareness of cooperation—seems to testify against the first formulation of the previously mentioned subjective connection. It is true that the second formulation is not covered by the joint participation regime, as it is based precisely on the ignorance (even if culpable) of such cooperation. Theoretically, there will therefore be room for such a requirement. We believe, however, that this requirement is out of place, since it would rarely respect the interests of the affected person. Except for the cases in which people are acting side by side, it would be almost impossible for an agent to be aware of the other (Staudinger and Eberl-Borges 2018, pp. 23 et seq, 35–36).

Likewise, there is no reason to limit the liability of agents based on their physical proximity, the temporal proximity of their conduct or the similarity between them. It is true that in some cases this will happen naturally. Consider, for example, those cases in which two people—not knowing each other—shoot at another, without being able to demonstrate which of the shots was fatal, insofar as any one of them could have been the cause of death. The contours of the situation show that

the subjects were necessarily physically close, that the corresponding actions were relatively synchronous and that they shared identical characteristics. Sometimes, however, this is not the case and there is no materially relevant reason to treat the problem differently, namely to deny protection to the victim's claims. This happens, for example, in those cases in which a person is infected with AIDS and it is not possible to determine, at the time the disease is detected, whether its origin is found in a contaminated blood transfusion that he had taken in the past or in intimate relations that he also had in the past with an infected person. If in both cases there is fault or negligence from the potential perpetrators, what are the grounds for rejecting the affected person's claim for compensation?

Holding the potentially harmful agents responsible implies placing the emphasis on the dangerousness of the action carried out by each of them (Bydlinski 1959, p. 13). It will not be the damage that justifies the liability of the agent—as he may not be its author–, but the ability of the action to produce it. The underlying logic seems to be more consistent with the idea that, in these cases, the offences should be seen as offences of concrete danger and not as offences of result.

Based on this, it is reasonable to hold that, whenever all the people involved in the AI systems individually practice an unlawful and culpable act, in abstract capable of producing the damage, the solution will be in principle the joint and several liability of those involved, even though it is not possible to identify the concrete action behind that damage. This means that each agent is liable for the totality of the damage and can then claim back the share that each one is liable for in internal relations.

If, in general, the need for an objective and / or subjective connection is very doubtful, in these situations we believe that such a requirement does not appear to make sense. The dispersion both in place of the subjects intervening in the process of creation and use of AI and in time, given the time that may mediate between those interventions, would hardly allow protecting the interests of the affected person. On the other hand, given the different nature of the involvements—creation, programming, insertion of data, updates, use of the AI system–, the affected person would hardly obtain compensation for the damage suffered. Although from a conceptual point of view this could be the solution, from a values point of view this is not the most appropriate outcome. In such a case, the important is to decide which of the interests deserves protection: the affected person's or the agents'. Bearing in mind that the latter have committed a fault that was able to cause the damage, there is no reason to give their interests primacy over the position of the affected person.

One of the problems addressed by the 2019 Report is precisely that of alternative causation. It is recommended that its regime should be similar to multiple causation, with any participant being jointly and severally liable for all damages suffered.[7] Although the actual cause of the damage is unknown, it may be possible to establish degrees of probability among the actions of the different agents. In such a scenario,

---

[7] Point 31.

it is recommended that the burden of proof be placed on the side of the person whose action has a higher probability of having caused the damage.[8]

It should be stressed that the 2019 Report proposes a fault-based liability system as the rule for civil liability, despite the fact that it admits a lightening of the rules on the burden of proof in matters of causality, taking into account:

1. "the likelihood that the technology at least contributed to the harm";
2. "the likelihood that the harm was caused either by the technology or by some other cause within the same sphere";
3. "the risk of a known defect within the technology, even though its actual causal impact is not self-evident";
4. "the degree of ex-post traceability and intelligibility of processes within the technology that may have contributed to the cause (informational asymmetry)";
5. "the degree of ex-post accessibility and comprehensibility of data collected and generated by the technology";
6. "the kind and degree of harm potentially and actually caused".[9]

"Where the damage is of a kind that safety rules were meant to avoid, failure to comply with such safety rules, including rules on cybersecurity, should lead to a reversal of the burden of proving:

(a) causation, and/or.
(b) fault, and/or.
(c) the existence of a defect".[10]

The 2020 EP Resolution does not address the problem of alternative causation. Although it establishes the joint liability of the various operators who may be held liable, it is not clear whether the rule is intended for cases of joint-participation, parallel authorship, alternative causation or for all.[11] This means that it is not certain the possible liability of the participants in the causal process when one can't determine which of the actions effectively caused the damage. Also, even if their liability is accepted, no position is taken as to the possible need to prove the suitability of the action for producing the damage.

According to this proposal, the basic liability system should be a fault-based liability, although it provides for a rebuttable presumption of fault from the operators.[12]

All in all, in cases of alternative causation the solution will necessarily be one of the following three:

(a) exclusion of civil liability;

---

[8] Points 25–26.
[9] Points 25–62.
[10] Point 24.
[11] Article 11.
[12] Article 8.

(b) partial liability of each participant for a share of the total damage;
(c) joint and several liability of all participants for the entire damage.

From a technical point of view, all of these solutions are viable. The first solution is grounded on the lack of concrete causation. The second and third, differently, place the emphasis either on the fault of each agent or on the damage suffered by the affected person. The only difference is the way the causal link is assessed. Instead of requiring proof of causation *in concreto*, causation *in abstracto* is sufficient. They just differ in the regime of compliance with the obligations imposed on them. In the second solution, the shared liability system is applied, that is to say, each agent is liable for only a portion of the compensation. The affected person cannot demand full compensation from a potential injurer, in the same way that none of the potential injurers is bound by the whole. In the third solution proposed, the regime of joint and several liability prevails, by means of which each agent is liable for the total compensation, with the possibility claiming the payment back in internal relations.

   The first solution grants more emphasis to the interests of the potential injurers in relation to the interest of the affected person. This does not appear to be the most appropriate one. The number of people involved in the creation and use of an AI system, located or coming from different areas of the globe and different fields of activity, makes it very difficult to identify and locate them. Therefore, it is too burdensome to impose on the affected person—often a natural person unaware of all these details—the need to sue each of the participants in order to obtain compensation for all the damage suffered. In fact, it will be less difficult for one of these participants to locate the others and exercise his right of claiming the payment back. For those reasons, the system of join and several liability seems more appropriate to the situation.

   A presumption of causal adequacy also seems appropriate in this context, given, on the one hand, the highly technical, specific and complex nature of the whole system and, on the other hand, the (not culpable) lack of knowledge of the potential victims as to how the system works.

   A final note to mention that the 'AI Liability Directive', while not addressing the problem of alternative causation, proposes two very important measures regarding the fault-based liability: the empowerment of national courts to order the providers or the users of the AI system to disclose relevant evidence at its disposal about a specific high-risk AI system (art. 3) and the establishment of rebuttable presumptions of the causal link between the fault of the defendant and the output produced by the AI system or the failure of the AI system to produce the output (art. 4).

3   Strict Liability

Quid iuris when no fault has been committed and the damage is due to the functioning of the AI-systems? The only possible path will be that of strict liability. In this context, two questions have been raised:

1. Are any of the strict liability regimes currently in force to be directly or by analogy applicable to the problem?
2. Is it necessary or advisable to design a specific regime for damages created by an AI system?

The current strict liability regimes that are presented as possible solutions to the problem are mainly: product liability, liability for damage caused by animals and liability for damage caused by a motor vehicle.

According to the Council Directive 85/374/EEC (PLD), the producer—understood to be the manufacturer or importer of goods into the EU for distribution as part of his commercial activity[13]—is liable for defects in his product.[14] Product means "all movables, with the exception of primary agricultural products and game, even though incorporated into another movable or into an immovable". Electricity is also considered to be a "product".[15] The producer is only liable for defects of the product at the time it was placed on the market and not for those that appear subsequently.[16] The victim is responsible for proving the damage, the defect and the causation of the damage by the defect.[17]

Several difficulties have been identified in applying this regime to damages caused by an AI system. First, this regime would not entirely solve the problem because it does not address the possible liability of the owner, holder or user of an AI system. This means that it could only offer a partial solution. Even in the field of the creation of AI systems, there are obstacles to its application (Barfield and Pagallo 2020, p. 96). It is the case with the definitions of producer and product. While there is no doubt that the manufacture of hardware can be seen as a production activity and the result as a product, the same is not true when it comes to creating the algorithms on which AI is based, or to feeding that system. The definition of product in the PLD may give the impression that only movable tangible things—i.e., things which can be perceived by the senses—deserve such a qualification. An algorithm or the data that feeds it can hardly fall into that category (Revolidis and Dahi 2018, p. 61; Capilli 2020, p. 478). It is therefore also difficult to regard a programmer or the person who feeds the data as producers within the meaning of the PLD. Of course, one could always try to see the norm as a living instrument subject to evolutionary interpretation, adjusting it to today's reality (and not to the standards of 1985), or, if this is not possible, resort to analogy (Wagner 2018, p. 11). However, it is uncertain whether this would be fruitful, since the producer is only liable for defects in the product which existed at the time it was placed on the market. The big problem with damages caused by AI systems lies in the fact that the risk of injury is more associated with the autonomy of these systems than with

---

[13] Article 3.

[14] Article 1.

[15] Article 2.

[16] Article 7.

[17] Article 4.

a possible defect in their design (Pagallo 2013, p. 117). In most cases, there is no defect. The system's evolution is not controllable by the designer, the programmer or the other people involved in feeding and updating it. Moreover, as a rule, errors occur long after the system has been placed on the market and were not known or were not identifiable at the time (Capilli 2020, pp. 459, 473–474; Molnár-Gábor 2020, pp. 253–254).

Facing these difficulties, the proposal for a new PLD establishes: an extension of the notion of product to explicitly include digital manufacturing files and software (art. 4), thus removing the uncertainty about the qualification of AI systems as a product; presumptions of defectiveness (art. 6); and presumptions of the causal link between the defect and the damage (art. 9). The empowerment of national courts to order the defendant to disclose relevant evidence at its disposal is also established (art. 8).

Strict liability is typically based on one of two pillars: the position of control of a source of danger or the taking advantage of that source of danger (Barbosa 2020, p. 40).

The liability for damage caused by animals seems to seek support precisely in the taking advantage of those animals by their owner. The possible application of this regime to damages due to AI systems would only be viable through analogy.

The similarity between the two cases is found in their unpredictability (Pagallo 2013, pp. 33, 38). Just like animals, the performance of the AI system is also unpredictable. And it is precisely this unpredictability that creates the risk underlying these two realities. From this perspective, nothing would prevent the application by analogy of the rules on liability for damage caused by animals to damage caused by AI systems.

The question is whether the proposed solution is the most adequate to the problem. If one looks at it, the option taken is to hold liable those who take advantage of the source of danger or those who take advantage in their own interest. This means that those who create the source of danger or those who take advantage of that source in the interest of others will not be liable for damages arising from it. Transposing this to the digital world, this is equivalent to exclude the liability of the creator of the AI-system and, if that's the case, of the user who uses it in the interest of others. In many cases this doesn't seem to be the most appropriate solution to our problem.

In fact, it should not be forgotten that those who design, program, feed and update these systems determine their functioning. In closed software systems, no one has access other than these entities. The degree of information and understanding of the system is also not at all the same as that of its users (Revolidis and Dahi 2018, p. 74). With this in mind, does it make sense to base liability totally and exclusively on the taking advantage of risk, excluding those who create it or who can limit it, to a greater or lesser extent? It could be said that the ultimate decision to use the AI system lies with its user. However, that decision does not mean controlling the risks of the system. This decision has no influence on the design of the AI model. It is important to distinguish between the intrinsic danger, resulting from the system's configuration, and the danger resulting from the decision to use that

system in inappropriate circumstances. In the first case, the danger comes from the system itself, in the second the danger results from a bad decision of its user. Here it is important to begin by asking if that bad decision constitutes sufficient ground for fault-based liability, particularly for violation of traffic duties. If this is not the case, we should rely on strict liability, which will be based on the taking advantage of the source of danger by the agent. There, on the contrary, the source of strict liability should be found in the dangerousness of the system, and it may be discussed whether without such a scenario it would be more appropriate to hold liable those who have a position of (relative) control of the source of danger or those who take advantage of that source or both.

We believe that the latter is the most adequate solution. It makes no sense to exempt from liability the designer of the algorithm, the programmer, the person who enters or updates data. They are in the best position to control this source of danger, and they also benefit from it, albeit indirectly (Wagner 2018, pp. 9–10). Although, as a rule, they do not benefit from the advantages created by the system, they take advantage of its value by trading it. Similarly, it is not reasonable to exonerate users from any possible liability. In addition to taking advantage of the source of danger, they themselves have the power to decide whether to use the AI system in those specific circumstances. Their decision to use it, while not being the exclusive cause of the danger, contributes to its maintenance or increase.

What should be determined is whether the damage corresponds to the materialization of one of the dangers generated or intensified by the creation and/or use of the AI system. In other words, the question is whether the damage is the result of the materialization of those dangers—of the system itself and of the decision to use it in that context and for that specific purpose—or only of one of them. If it comes from both, there should be joint and several liability of all participants in the causal process. If it comes from only one of them, he alone should be liable. However, we must pay attention to the fact that within each group—creators or users—there may be several potentially harmful persons. In the impossibility of determining the dangerousness of each individual participation and the contribution of each to the production of the damage, each potentially harmful person should answer jointly and severally for the damage (Ebers 2016, p. 16; Capilli 2020, p. 477).

The liability regime for damage caused by animals does not appear to cover all these situations.

The liability regime for damage caused by motor vehicles doesn't seem a solution to our problem either, since it would once again penalize the user of the AI system. In fact, the liability for the damages caused by a motor vehicle always lies with the owner or user of the vehicles, as he is the one who benefits from it. For the reasons already mentioned, such a vision would not be the most appropriate answer to our problem.

Some legal systems however demand not only that the responsible person use the vehicle in his own interest, but also that he or she was actually driving, suggesting the need to have a position of control over de source of danger. Such a regime discards the owner or users' liability, when they use the vehicle in the interest of a third party, and the potentially liability of the system designers.

It is important to understand, however, that this position of control concerns only the possibility to determine whether the vehicle is used and how it is used. No control would be required over the proper construction and performance of the vehicle. From a subjective point of view, this is important because it excludes the manufacturers of such vehicles from the scope of application of this regime. This means that the designer of the algorithm, the programmer and the people who feed or update the system are also excluded from liability here. Only the owner, the holder or the user remain liable.

The imperfection of the machine justifies a strict liability system. This imperfection also exists in AI systems. Therefore, it is also possible to apply analogy to damage caused by AI systems. In some cases, it may not be necessary to use analogy and the rules in question may even be directly applicable to the situations. This happens, for example, in accidents involving autonomous vehicles, although it is questionable to what extent there will be effective direction of the vehicle in cases of full automation (Barbosa 2020, p. 286).

Nevertheless, we have doubts as to the adequacy of theses regimes to solve our basic problem, since it once again penalizes the user of the AI system in his own interest, discarding his liability in the hypotheses of use in the interest of a third party and, more importantly, excluding the potentially liability of the system designers.

VI. The 2019 Report supports the adoption of strict liability for operators benefiting from or controlling the system. They limit such liability, however, to cases where AI-systems are used "in non-private environments" and "may typically cause significant harm".[18] This therefore excludes cases where the system is used in a closed environment, exposing a small number of people to the risk of injury, which can happen, for example, in the use of AI in the performance of a medical procedure. The possible extent of the injury appears to outweigh its gravity. If there is an operator who benefits from the risk and another who controls it, "strict liability should lie with the one who has more control over the risks of the operation", thereby showing the primacy of control in relation to the position of profiting from the source of danger.[19] The report also advocates extending the product's liability to cover defects in software that occur after it have been placed on the market.[20]

The 2020 EP Resolution limits the strict liability of operators to damage caused by high-risk AI-systems, listed in the annex to the proposal. According to the proposed regime, high-risk AI-systems should be understood as having "a significant potential in an autonomously operating AI-system to cause harm or damage to one or more persons in a manner that is random and goes beyond what can reasonably be expected; the significance of the potential depends on the interplay between the severity of possible harm or damage, the degree of autonomy

---

[18] Point 9.

[19] Point 11.

[20] Points 13–14.

of decision-making, the likelihood that the risk materializes and the manner and the context in which the AI-system is being used".[21]

Common to these proposals is the idea of trying to limit strict liability to certain cases. The justification for this lies not so much in legal considerations as in policy. The aim is to establish a regime that does not discourage AI-systems scientists and developers from continuing their research and activities. The choice is understandable, although it is difficult to accept the results to which it leads. It is inconceivable, for example, that a patient who has suffered serious damage to his or her life or physical integrity as a result of a medical procedure using an AI-system would not be compensated. A judgement of proportionality and reasonableness prevents the violation of any good equal or superior to the good being protected. It is unlikely that technological development will be a superior good to life or even, in certain cases, to physical integrity. Moreover, the fact that developers are not held responsible does not encourage them to invest their resources in improving the system (Wagner 2018, p. 18).

A compensation fund or a compulsory insurance will only prevent this inconvenience in the event that there is liability for damages suffered by someone as a result of the action of an AI-System, regardless of whether fault or strict liability is involved. The simple proof of the damage—although it may make sense to limit the damage compensable by this fund according to its nature and gravity—and respective cause would be sufficient to justify the compensation supported by the fund or the insurer. Otherwise, the fund or insurer would only substitute the injurer in fulfilling his obligation to compensate, without extending the protection of the interests of potential injured parties.

Based on all these thoughts, we advocate for a liability regime that deals precisely with these problems. Otherwise, on one hand, many situations will remain unprotected and, on the other hand, users of this type of systems will be held liable above all, exonerating the developers from any liability. As already mentioned, this does not seem appropriate.

Unfortunately, the Draft AI Liability Directive initial proposal does not accept a strict liability approach. Instead, it accepts a fault based liability, with some specific tools: a rebuttable presumption of causality and a disclosure of evidence regime. For the reasons explained before, we do not think that the proposed regime is adequate to deal with damages caused by AI systems. In addition, it is subject to question if the Draft Directive is consistent with the level of protection envisaged by the Draft AI Act.

---

[21] Article 3.

4 Exemption from Liability for Damage Caused by an AI System

Depending on the nature of the liability, another question should be posed: in which situations can the agent escape liability? We seek here to address the cases in which the agent's liability should be excluded.

We let outside the scope of this analysis the factors that can exempt a producer from liability and if the factors set out in article 7 PLD need to be revisited, as they are inadequate to address the specificities arising from damage caused by AI systems. We acknowledge that there is already relevant and reasonable doctrine which underlines that the directive allows an AI producer to avoid liability by invoking the so-called *development risk defence* (Bertolini 2020, p. 58; Evas 2020, p. 9; Navas 2020, pp. 80–81). A concern that the proposal for a new PLD seems to have addressed since the development risk defence cannot be invoked when the scientific and technical knowledge evolution occurs in the period in which the product was still within the manufacturer's control. [22] This amendment suggests that the manufacturer remains responsible if, for example through updates, he/she can eliminate the defects revealed by the evolution of knowledge and technique. However, it must be questioned whether such change is sufficient to ensure the safety of AI-enabled products already put into circulation.

Therefore, we will focus on the applicable exclusions when liability is not based on the AI system's defects.

Naturally, there is no unitary answer to the question, since the grounds to exempt an agent vary due to the different nature of the liability.

If the agent is liable based on strict liability, the grounds to escape liability will differ from those that should be accepted when we have a fault-based liability regime, even with a reversal of the burden of proof. For the present purpose, it is less important determining whether the agent is a programmer or a user, a backend operator or a frontend operator, as defined above, than to understand whether he/she is liable based on strict liability system or on fault-based liability regime.

Actually, both in member states' tort legislations (see Evas 2020, pp. 10–33)[23] and in EU proposals to harmonize a tort law regime for damage caused by an AI system, there is a trend to exclude "one fits all solution", which means that the obligation to compensate damages caused by IA systems may be based either on strict liability or fault-based liability.

Nonetheless, the European Commission's proposal for an AI Liability Directive only addresses the harmonization of the rules for the presumption of causality and the disclosure of evidence, leaving it up to each Member State to determine whether agents' liability should be based on strict liability or a fault-based liability system.

---

[22] Article 10/1/e).

[23] For a detailed analysis of the liability in national law of each member states and its possible application to AI.

If the agent is going to be held liable for damage caused by an AI system, he cannot dismiss liability because the damage was caused by that system or is a consequence of its autonomy.[24] However, in theory, several other factors can exempt an agent from liability:

(i) the proof that the agent complied with specific duties, such as, for example, diligence, custody and surveillance, and acted with due care;
(ii) the proof that harm or damage was caused by force majeure;
(iii) the proof that harm or damage is attributable to a third party;
(iv) the proof that the victim or the affected person caused harm or damage;

The first factor can only be admitted in a fault-based system, even with a reversal of proof, since, in cases of strict liability, the agent's liability is not based on the existence of an unlawful act or a breach of a duty.

In member states' tort laws, the liability regimes based on a presumption of fault allow agents to escape liability, when they prove to have acted with due care to avoid damage. The proposals drawn up by the Expert Group and the EP[25] also seem to accept this solution, seeking to tailor the evidence that should be produced to address the peculiarities arising from damage caused by AI systems. In this light, the Expert Group proposes: "Operators of emerging digital technologies should have to comply with a*n adapted range of duties of care*".[26]

In a strict liability system, the agent will be liable regardless of having breached the incumbent duty. His/her liability will be justified by the risk that the agent generates with the development of his/her activity or by the profits from which he/she benefits.

Regarding damages caused by force majeure or a fortuitous event, there is no doubt that the demonstration of its existence should lead, in principle, to the exclusion of the agent's liability, either in a system of presumption of fault or in a system of strict liability (Pagallo 2013, p. 33). Nonetheless, some clarifications are in order.

First, even in a fault system, if it is proved that the damage was directly caused by force majeure, but at the same time, it is shown that, if the agent had acted diligently, he could have avoided the damage, the agent's liability should remain and apply.

Secondly, in strict liability regimes, we tend to consider that the proof of the existence of force majeure that caused the damage is sufficient to remove the agent's liability. The comprehensive formulation that is sometimes used should be reconsidered, from our point of view. In articles 4 and 8 of the EP 2020 Resolution, it is proposed that *"the operator shall not be liable if the harm or damage was caused by force majeure".*

---

[24] Articles 4 (3) and 8 (2) EP 2020 Resolution.

[25] Article 8 EP 2020 Resolution.

[26] Point 16.

In order to escape liability, it should not be sufficient to prove the existence of force majeure. It should also be necessary to demonstrate that the force majeure is "alien" to the operation of the AI system.

Consider two examples: should the agent be liable for damage caused by a surgeon robot that, following an earthquake, falls over a patient injuring him/her? Should the agent be liable for damage driven by a surgeon robot that causes injury to a patient due to a connection failure caused by a severe storm?

In the first case, the damage caused by the robot could have been originated from any other instrument present in the operating room. The same cannot be said of damage deriving from a lack of connection, even due to an exceptional atmospheric phenomenon.

It is very doubtful that, when the agent is strictly liable, his/her liability can be removed when the force majeure event is not foreign to the functioning of the AI system. In fact, one of the characteristic risks of AI systems is that they may cause damages to third parties, when there is a connectivity failure and, therefore, even if that failure is due to an exceptional or unusual situation, the agent should not escape liability.

We are fully aware that resorting to an indeterminate or vague concept such as damages caused by force majeure alien or foreign to the utilization of the AI system will require an increased effort for the courts. Even if the situations in which there is a connection failure may be easier to understand and frame, practice and day-to-day events will certainly bring other examples that will certainly raise more questions.

Another question frequently arising is how to deal with cases where the agent can prove that a third party caused the harm or damage. The autonomy of this question— and ultimately the exclusion of liability - depends on whether the damage has been caused exclusively by a third party who is not a producer or an operator of the AI system. Previously, we have already dealt with the problems arising from hypotheses in which several operators may be liable based on a strict liability system or on a fault-based liability regime.

The issue here is to identify and segregate the cases in which the third party interfered with the AI system, modifying, or affecting its operation. We can identify several examples. The hackers who maliciously interfere with the AI system. The subjects who negligently disrupt the functioning of a robot by disconnecting its power supply. The children who hijack a goods delivery drone, etc.

In any of the situations described, it is unquestionable that the third party should respond for the damages suffered by the injured party. The problem is to ascertain if this third-party liability can exclude the liability of a producer or an operator, as defined above.

As mentioned before, we will exclude from our analysis the producer's or the manufacturer's liability for defective products.

In a negligence-based liability system with presumed fault, the proof that a third party exclusively caused an injury should in general allow the agent to escape liability, unless the third party's action was enabled by the agent's breach of due diligence. In other words, the proof that a third party exclusively caused the damage

should exclude the agent's liability, except in cases where the agent could have prevented the third party's action if he had acted diligently.

As regards strict liability, there is no uniform solution for damage exclusively caused by third parties. The rule is that strict liability should be excluded in these cases, although joint and several liability is admitted in specific regimes. The most paradigmatic example is vicarious liability, when it is set that a principal is vicariously and strictly liable for the torts of his/ her agents.

The absence of a uniform system for dealing with damage caused exclusively by third parties is an argument in favour of an autonomous regime for damage caused by AI systems.

The EP 2020 Resolution is very innovative on this point. According to Article 8.3, "where the harm or damage was caused by a third party that interfered with the AI-system by modifying its functioning or its effects, the operator shall, nonetheless, be liable for the payment of compensation if such third party is untraceable or impecunious". Assuming that it is often difficult to identify the person of the third party and/or that the third party may not have sufficient assets to pay the compensation, it is proposed that the agent's liability be maintained, even in cases framed under the fault-based liability regime. On the other hand, this option implies that, *a fortiori*, the solution should apply to the hypotheses of strict liability.

Although we understand the concern behind the proposal, it is difficult not to question whether we are not facing strict liability (Antunes 2020, p. 10), in spite of the qualification proposed by the European Parliament. When it is proven that the agent acted diligently and could not have avoided the third party's action and, in spite of this, he/she is still held liable since the party is untraceable or impecunious, the conclusion can only be that the EP 2020 Resolution favours a strict liability approach.

When an AI system causes damage, one can never set aside the possibility that the affected person has by his/her action or omission contributed to the damage suffered or to its extent. According to a more modern understanding, the agent's liability should only be excluded when the behaviour of the affected person is the sole cause of the damage. In other cases, the negligent conduct of the injured party should only constitute grounds for reducing liability. This solution is embraced by the 2019 Report[27] and the EP 2020 Resolution.[28]

There are, however, legal systems, such as the Portuguese, where the Civil Code still provides for a total exclusion of liability in some of the cases described (cf. article 570, no. 2). The solution has, however, been frequently criticized and is inadequate for situations in which the damage was simultaneously caused by the AI system and by the injured party. This is just another example of the need of an autonomous civil liability regime for AI damage. [29]

---

[27] Point 28.

[28] Article 10.

[29] *See* also, on the AI Act, in this book P U Lima and A Paiva - Autonomous and Intelligent Robots - Social, Legal and Ethical Issues; M N Duffourc and D S Giovanniello - The Autonomous

References

Antunes HS (2020) Civil liability applicable to artificial intelligence: a preliminary critique of the European Parliament resolution of 2020. Available via SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3743242. Accessed 18 Apr 2022

Barbosa MM (2020) O futuro da responsabilidade civil desafiada pela inteligência artificial: as dificuldades dos modelos tradicionais e caminhos de solução. Revista de Direito da Responsabilidade ano 2:280–326

Barfield W, Pagallo U (2020) Advanced introduction to law and artificial intelligence. Edward Elgar Publishing, Cheltenham

Bertolini A (2020) Artificial intelligence and civil liability, European Parliament's Committee on Legal Affairs. Available at https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf. Accessed 18 Apr 2022

Brambring G (1973) Mittäter, Nebentäter, Beteiligte und die Verteilung des Schadens bei Mitverschulden des Geschädigten. Dunker & Humbolt, Berlin

Bydlinski F (1959) Haftung bei alternativer Kausalität. Jurist Bl 81:1–13

Capilli G (2020) I criteri di interpretazione dele responsabiitá. In: Alpa G (ed) Diritto e intelligenza artificiale. Pacini Giuridica, Pisa, pp 457–487

Chesterman S (2021) We, the robots? Regulating artificial intelligence and the limits of the law. Cambridge University Press, Cambridge

Ebers M (2016) La utilización de agentes electrónicos inteligentes en el tráfico jurídico: necesitamos reglas especiales en el derecho de la responsabilidad civil? In: InDret - revista para el anàlisis del derecho, 3. Universitat Pompeu Fabra, Barcelona, pp 1–22, available at http://www.raco.cat/index.php/InDret/article/view/314400. Accessed 18 Apr 2022

Ebers M (2020) Regulating AI and robotics: ethical and legal challenges. In: Ebers M, Navas S (eds) Algorithms and law. Cambridge University Press, Cambridge, pp 37–99

European Commission (2021) 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, COM/2021/206 final

European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)), OJ C 404, 6.10.2021, pp 107–128

Evas T (2020) Civil liability regime for artificial intelligence - European added value assessment, European Parliamentary Research Service, Available at https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654178/EPRS_STU(2020)654178_EN.pdf, Accessed 18 Apr 2022

Expert Group on Liability and New Technologies – New Technologies Formation, Liability for artificial intelligence and other emerging technologies, Directorate-General for Justice and Consumers (European Commission) (2019). Available at https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-1aa75ed71a1/language-en. Accessed 18 Apr 2022

Larenz K (1994) Lehrbuch des Schudrechts, 2. Bd. 2. HB, 13. edn. C.H. Beck'sche VerlagsBuchhandlung, München

---

AI Physician - Medical Ethics and Legal Liability; D Durães, P M Freitas and P Novais - The Relevance of Deepfakes in the Administration of Criminal Justice; A Keller, C Martins Pereira and M Lucas Pires - The European Union's approach to Artificial Intelligence and the Challenge of Financial Systemic Risk; and J C Abreu - The "Artificial Intelligence Act" Proposal on European e-Justice Domains Through the Lens of User-focused, User-friendly and Effective Judicial Protection Principles. At the same time, cfr in this book M Lanz and S Mijic - Risks associated with the use of natural language generation - Swiss civil liability law perspective.

Molnár-Gábor F (2020) Artificial intelligence in healthcare: doctors, patients and liabilities. In: Wischmeyer T, Rademacher T (eds) Regulating artificial intelligence. Springer, Cham, pp 339–360

Navas S (2020) Producer liability for AI-based technologies in the European Union. Int Law Res 9:77–84

Pagallo U (2013) The laws of robots – crimes, contracts, and torts. Volume 10 of law, governance and technology series. Springer, Dordrecht

Revolidis J, Dahi A (2018) The peculiar case of the mushroom picking robot: extra-contractual liability in robotics. In: Corrales M, Fenwick M, Forgó N (eds) Robotics, AI and the future of law. Springer, Singapore, pp 57–79

Staudinger J, Eberl-Borges C (2018) Kommentar zum bürgerlichen Gesetzbuch mit Einführungsgesetz und Nebengesetzen, 2. Buch – Recht der Schuldverhältnisse, §§ 830-838. Sellier – de Gruyter, Berlin

Turner J (2019) Robot rules- regulating artificial intelligence. Palgrave Macmillan, London

Wagner G (2018) Robot liability. Available via SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3198764. Accessed 18 Apr 2022

Wagner G (2020) Münchener Kommentar zum Bürgerlichen Gesetzbuch. C.H. Beck, München