Suggestions for a Revision of the European Smart Robot Liability Regime

Chiara Gallese
University of Trieste, Trieste, Italy
Carlo Cattaneo University, Varese, Italy
Eindhoven University of Technology, Eindhoven, The Netherlands
cgallese@liuc.it

Abstract: In recent years, the need for regulation of robots and Artificial Intelligence, together with the urgency of reshaping the civil liability framework, has become apparent in Europe. Although the matter of civil liability has been the subject of many studies and resolutions, multiple attempts to harmonize EU tort law have been unsuccessful so far, and only the liability of producers for defective products has been harmonized so far. In 2021, by publishing the AI Act proposal, the European Commission reached the goal to regulate AI at the European level, classifying smart robots as "high-risk systems". This new piece of legislation, albeit tackling important issues, does not focus on liability rules. However, regulating the responsibility of developers and manufacturers of robots and AI systems, in order to avoid a fragmented legal framework across the EU and an uneven application of liability rules in each Member State, is still an important issue that raises many concerns in the industry sector. In particular, deep learning techniques need to be carefully regulated, as they challenge the traditional liability paradigm: it is often not possible to know the reason behind the output given by those models, and neither the programmer nor the manufacturer is able to predict the AI behavior. For this reason, some authors have argued that we need to take liability away from producers and programmers when robots are capable of acting autonomously from their original design, while others have proposed a strict liability regime. This article explores liability issues about AI and robots with regards to users, producers, and programmers, especially when the use of machine learning techniques is involved, and suggests some regulatory solutions for European lawmakers.

Keywords: robot law, liability, AI

1. Regulating robots in the EU: The European approach

In recent years, due to the increasing use of new technologies, the EU has tried to regulate robots and AI in order to harmonize the legal framework within the Member States.

On 16 February 2017, the European Parliament adopted a Resolution on Civil Law Rules on Robotics, where it asked the Commission for a proposal to amend the civil law rules on the liability of robots and AI. On 25 May 2018, the European Commission adopted a Communication on Artificial Intelligence for Europe, addressing the new challenges related to AI, issued a working document on liability for emerging digital technologies, and, in December 2018, it published a coordinated Plan on Artificial Intelligence. In April 2019, the High-Level Expert Group on Artificial Intelligence published ethics guidelines for trustworthy Artificial Intelligence.

The matter of liability has been thoroughly examined in the report of the Expert Group on Liability and New Technologies – New Technologies Formation (2019), which suggested two different regimes of civil liability depending on the degree of risk, that is strict liability for high-risk cases, and subjective liability with a presumption of fault in all the others. This model has been discussed and welcomed by a number of scholars (Karner, 2019; Spindler, 2019; Barbosa, 2020).

Later that year, a new resolution was adopted (EU Parliament, 2020), following a different route, but maintaining the dual liability distinction. At the end of the document, the text for a proposal is attached. As noted by Sousa Antunes (2021), the proposal has many flaws, for example, "it is difficult to understand why the European Parliament made compensation for it dependent on an economic repercussion of the offence".

In addition, a maximum amount that limits the compensation for harm caused by high-risk systems has been established, following Article 16 of Directive 85/374/EEC. Sousa Antunes suggests that "It would, perhaps, make more sense to set incremental limits according to the foreseeable impact of the harm calculated by the number of people harmed. In Portuguese law, for example, the maximum limits of compensation in the case of traffic accidents differ depending on whether the transport is individual or collective".

2. Machine Learning and black-boxes: A new challenge for liability rules

From a legal perspective, particular attention should be paid to models based on machine learning (especially deep learning) and those based on a black-box approach. In fact, in these cases, programmers create the model and provide relevant examples, but they do not know the final outcome, as the model is designed to learn by itself (prior to its commercialization, during the training phase, but also after it is put in the market). This often means that it is not possible to know the reason behind the output given by those models, therefore it is important to explore and understand how they behave in different feature subspaces (Lakkaraju et al., 2019).

The second element to consider when exploring the possibilities of regulating smart robots is that it is not possible to a priori exclude the presence of biases or bugs (Naur and Randell, 1968) that could influence the behavior or the decisions taken by the robot. In fact, the model is developed by a human being, who is, of course, fallible. This means that a programmer will never be able to ensure that the model behaves exactly as programmed, but it also means that, by changing hyper-parameters and selecting the data conveniently (or politically), the same tool can be used to obtain different results.

The data utilized can, in itself, represent a relevant problem: they can be wrong, false, incomplete, or simply too few to be meaningful, therefore the result will not be reliable; moreover, even if the data are correct, artificial intelligence can reach conclusions that are prejudicial to certain categories, based on dynamics deriving from inequalities present in society, which, in the end, can confuse the relationship between cause and effect, as shown in some famous cases (Falletti, 2020).

This poses a great legal problem from multiple points of view (Marchant and Lindor, 2012), as it challenges the traditional liability paradigm: is it possible to adapt a strict liability regime to a situation in which nor the programmer nor the manufacturer are able to predict the AI behavior? And from a technical point of view, is it safe to sell to customers a product that, even if trained by the producer, cannot be fully explained? Some have theorized a so-called responsibility gap (Matthias, 2004), while others have opposed this view (Tigard, 2020). According to Tigard, the recent literature is polarized between “techno-optimists” (e.g., Santoro et al., 2008; Hanson 2009; Rahwan 2018; Nyholm 2018) and “techno-pessimists” (e.g., Sharkey, 2010; Asaro, 2012; Char et al., 2018; Danaher, 2016). Some authors believe that the current legal system is adequate to regulate machine learning liability (Amidei, 2019).

The subject who is held responsible for the use of AI may vary depending on the situation and the different Member States regulate the matter differently. Some authors have argued that we need to take liability away from producers and programmers when robots are capable of acting autonomously from their original design (Marchisio, 2020; Scherer, 2015) because it would be virtually impossible to foresee any possible situation that could occur (Zipp, 2016). To some extent, this could be true, but there are many issues to consider before accepting this solution.

As noted by the 2020 resolution, “fair compensation procedures mean that each person who suffers harm caused by AI-systems or whose property damage is caused by AI-systems should have the same level of protection compared to cases without involvement of an AI-system”. This principle means that, from the victim’s perspective, it is not relevant whether the harm has been caused by an autonomous system or by a human: they are still entitled to the same compensation.

3. Autonomous liability for autonomous robots?

So far, there are no AI models that can reach human (or even animal-like) intelligence, and probably never will. It is, therefore, improper to implement a legal framework in which smart robots have their own legal entity (Kamyshansky et al., 2019) and liability – especially criminal liability (Hallevy, 2010) – even if this could be theorized with a view for future development(Sousa Antunes, 2021). If we think about the original ratio of the provisions introducing civil liability, which is to provide compensation in case someone’s interest is damaged and to deter unlawful behavior, it is easy to see that, to this day, a robot could never be able to fully compensate a human being by itself, not until it will have its assets, its wage or any other way that a human being (or another legal entity) can find to earn money independently. In fact, it is possible to have an autonomous legal status for companies and associations only because they are composed of humans behind them. Not to mention that, so far, a robot is not able to fix bugs and patch itself, therefore, not only it would

be impossible to force it to comply with regulations about mandatory updates, but also, in case of a security breach of an outdated robot, there would be no one that could be considered liable for negligence.

Providing compulsory insurance is not enough to ensure the legal certainty that is required by liability systems, otherwise, following the same principle, it would be possible to cancel the civil liability of any company just enacting a law that requires all legal persons to have adequate insurance coverage, without any guarantees of solvency. In fact, moving the solvency issue from a State law system (such as the civil code provision on liability) to a contractual regime (such as a compulsory insurance system) would endanger the protection of citizens' relevant legal interests.

In conclusion, an autonomous legal status for robots is not desirable for any legal system. Instead, human or corporate liability should always be present.

4. The difficult scenario of open source robots

A different case concerns open-source software (Bahn and Dressel, 2006). This subject differs greatly depending on the legal system, as it involves copyright issues (who owns the final software after it has been modified by multiple programmers independently or collectively?) which are differently regulated by civil law and common law. As noted by Calo, the legal differences between closed-robot and open source-robots involve architectural differences between the two (Calo et al., 2016), both in terms of foreseeability of possible liability risks and control over intended use or functionalities.

Usually, the first developer is a single programmer who wants to publish his or her creation, hoping to help other programmers, and the first version of the software is uploaded in a common environment where others will add their versions. However, users can take the first version and develop other features on their own, without sharing their versions. The changes to the original version can be infinite and might relate to the behavior, interface, file format, data and parameters, the functioning of different parts, optimization, external resources, and many others, and it is a common practice to completely restructure the architecture of a software program (preserving the functionality) to improve performances or maintainability (refactoring). In general, the main developer has control over these integrations of the source code, and can decide whether to "merge" or refuse them; however, any developer can, in principle, take the original source code and "fork" a new version of the software, where the original creator has no longer control.

In this context, it is very difficult to regulate the liability of the software programmers, even when the license contains clauses excluding the liability since those clauses are invalid in many legal systems (see, for instance, art. 1129 of the Italian Civil Code, stating that no clause can exclude liability arising from wilful misconduct or gross negligence) and many judges do not look at them with favor.

To understand the liability arising from an open-source software the first issue to solve is to understand if the license is a contract or not, and the doctrine has no common ground on this issue; in fact, some authors believe the license to be a unilateral act (Piana, 2006). However, we believe that it can be attributable to the "general term and conditions" situation (art. 1341 of the Italian Civil code), a contract that is accepted by whomever downloads, uses or modifies the software program.

The second step is to understand if it is possible to consider open source software as "products" that are "put into circulation" according to the Product Liability Directive (Alheit, 2001) even when they are not released on the market (that is, for example, a software store) for a specific purpose but they are just uploaded to a public repository (such as GitHub).

In the O'Byrne case, the Court of Justice stated that "[...] a product must be considered as having been put into circulation, within the meaning of Article 11 of the Directive, when it leaves the production process operated by the producer and enters a marketing process in the form in which it is offered to the public in order to be used or consumed". However, the very nature of an open source code is to be later modified, corrected and improved (even by other users), and cannot always be considered as a finished product that has already "left the production process", let alone entering "a marketing process". If the applicability of the Directive to closed-source software is disputed, greater doubts arise regarding its applicability to open-source software.

5. Suggestions for a revision of the current liability regime

The liability regime in the EU needs to be updated to meet the new challenges derived from technological progress. For example, Article 7 of the European Product Liability Directive states that producers shall not be liable if it is proven that "having regard to the circumstances, it is probable that the defect which caused the damage did not exist at the time when the product was put into circulation by him or that this defect came into being afterwards", and "that the state of scientific and technical knowledge at the time when he put the product into circulation was not such as to enable the existence of the defect to be discovered". Those principles, however, are dangerous in the context of an AI based on a black-box model, especially if it is still learning after the release into the market, a scenario that is already possible since continual lifelong learning with neural networks already exists (Chen and Liu, 2018; Parisi et al., 2019; Mi et al., 2020).

The definition of "defect" itself is quite problematic; in fact, such a product would be, by design, "incomplete", even if based on state-of-the-art knowledge, and the "defects" that could cause harm could be introduced after the product was already sold. A simple bug could not be considered a defect, or it would be impossible to sell any software at all because, as we noted before, it is not possible to guarantee that a piece of software has no bugs; the notion should be much more restrictive and include only errors that are foreseeable by a reasonable programmer having in mind the standard knowledge in the field and the circumstances.

In addition to this problem, since Article 4 requires the injured person to prove "the damage, the defect and the causal relationship between defect and damage", with a black-box model it would be very difficult to understand why the AI behaved in a certain way and if the model was indeed "defective", thus making it difficult to prove the causal relationship between the model implementation and the damage. Existing a scientific consensus on the fact that the outcome of certain models based on machine learning are unpredictable, and in absence of a ban on such a technology to be released into the market, or of the mandatory presence of a log system, the Product Liability Directive should be considered inadequate to protect users.

Policymakers, after having implemented all relevant standards and technical regulations to ensure that only safe and secure products are released to the market, should carefully assess in what cases lifelong learning models should be allowed to the market, and, where appropriate, leave part of technological progress as a prerogative of scientific research and public institutions monitored by the Member States at a central level, in particular in case of dangerous products. Nevertheless, while addressing the distribution of liability issue, the difference between individual users and professionals and the unbalance of contractual power between each of them has to be considered. While it would be less problematic to shift the liability burden from producers to professionals (such as in the case of business-to-business relationships), lawmakers should be very careful in considering the release of producers from liability when it comes to individual users.

One of the biggest issues that prevent the EU from implementing an effective European discipline is the differences among tort systems within the EU. It would be useful to further research the liability systems of different Member States from a comparative perspective, in a way that a harmonized solution can be reached. The notions of fault, damage, causation, compensation, tort, and obligation should be explored in order to shape a common liability framework that can balance all the relevant interests in light of the peculiarities of smart robots. Product Liability Directive should be revised to adapt to new technological challenges, taking into consideration the peculiarities of machine learning and open-source software.

In reviewing the existing discipline, it would be more efficient to follow a case-by-case approach, modulating the liability system based on different situations: i) no-fault (when users' fundamental rights have to be protected from any harm) ii) no-fault unless specific circumstances arise (i.e., fortuitous event, or when the damage is caused by users or third parties) iii) fault only in the case of wilful misconduct or gross negligence (when certain standards are met and in case of open source license).

For machine learning-based robots, a specific regulation at the European level is needed since existing rules are not adequate to regulate all the issues that could occur after the release into the market of such devices. In particular, considering the variety of such technologies, each sector should have distinct rules. The costs and risks deriving from the delay in adopting new regulations in response to market requests should be taken into account as well. When the AI can learn by itself after the release on the market, if standards and safety

regulations are met, the *nullum crimen sine culpa* principle could be followed for the producer, with the exception of Constitutional rights breaches, thus regardless of the user acting as a consumer or a professional, and with the exception of producer's and programmer's wilful misconduct or gross negligence. In the case of a black box, however, it is not possible to dodge the producer's liability, not only because the users will never be able to understand and explain the outputs received from the device, but also because it poses security risks (Oh et al., 2019).

Regarding open source software, we believe that since most of the time it is impossible to trace back to a single programmer the causal relationship between the specific software implementation that caused damage and the damage itself, and since the software is published for free as a service to the community the liability of programmers should be limited to willful misconduct and gross negligence. Calo advocates against a blanket immunity to open-robot manufacturers (similar to that previously extended to the aviation industry in the nineties) and prefers to apply immunity only "in those instances where it is clear that the robot was under the control of the consumer, a third party software, or otherwise the result of end-user modification" (Calo et al., 2016).

6. Types of damages that can be compensated

A final remark should be made with regard to the types of damages that can be compensated. We think that the so-called "real notion of damage" and the so-called *Differenztheorie* are not adequate to compensate the damage arising from the use of a smart robot, as the economic rationale is not the only criterion that should be taken into account: the economic value of the robot itself is certainly not enough to cover all the possible damages, nor is the asset differential before and after the damage occurred. The material damage is not the only kind of damage that should be compensated.

On the other hand, we have seen that, under the product liability regime, economic loss or infringement of individuals' rights are not compensated. This means that, when the damage is caused by the mechanical part of the robot, users will not be adequately compensated. This system should be revised in light of the particular nature of smart robots, which surely could cause individuals' right infringement and economic loss on many occasions. In implementing a new regulatory system, the fact that the damage results from the breach of manufacturers, developers, programmers, service providers as well as backend operators; calls on the Commission to consider reversing the rules governing the burden of proof for harm caused by emerging digital technologies in clearly defined cases, and after a proper assessment; points out the importance of ensuring that the updated Union act remains limited to clearly identified problems for which feasible solutions already exist and at the same time allows future technological developments to be covered, including developments based on free and open source software; notes that the PLD should continue to be used with regard to civil liability claims against the producer of a defective AI-system, when the AI-system qualifies as a product under that Directive".

The 2020 resolution recognizes that some Member States have adequate liability rules: "there is no need for a complete revision of the well-functioning liability regimes, but [...] the complexity, connectivity, opacity, vulnerability, the capacity of being modified through updates, the capacity for self-learning and the potential autonomy of AI systems, as well as the multitude of actors involved represent nevertheless a significant challenge to the effectiveness of Union and national liability framework provisions", and "the existing fault-based tort law of the Member States offers in most cases a sufficient level of protection for persons that suffer harm caused by an interfering third party like a hacker or for persons whose property is damaged by such a third party, as the interference regularly constitutes a fault-based action; notes that only for specific cases, including those where the third party is untraceable or impecunious, does the addition of liability rules to complement existing national tort law seem necessary". The last sentence is arguable: the fact that some Member States have a good liability regime in most situations does not imply that the whole EU would not benefit from a harmonized framework of tort law.

In fact, we believe that moral, existential, and biological damages should be compensated too and this should be clearly stated in the new regulations, but without restricting the type of non-economic damages that can be compensated.

We argue, as well, that a smart robot can cause emotional distress, as research shows that humans display emotional reactions towards robots (Kwak et al., 2013; Rosenthal-von der Putten et al., 2013). As robots will be more and more present in our everyday life, even assisting the elderly, children, and people with disabilities, it is not unlikely to expect that humans may get attached to them and be hurt by them. This kind of distress should be evaluated by the judge and fairly compensated.

The remedies to compensate the damages should also be different from the sole economic reparation: for example, if a robot publishes a private conversation between two people, and one of them wants it to be removed because it is detrimental to their reputation, the judge should be able to order to the producer of the robot or to its owner to delete the conversation and to publish an apology statement instead. This remedy may be relevant in the case of cyberbullying, a problem that led many teenagers to suicide: if the robot is capable of learning from experience, then it can be trained (intentionally or not) to behave in an inappropriate way. Forcing a robot to say "I'm sorry" and to refute the hurtful words it said or wrote may look like a laughable solution to adults, but, in many cases, it could be important to kids, who often perceive their inanimate objects – such as virtual assistants – as real persons or living being (Beran et al., 2011).

7. Conclusion

The landscape of liability in the EU is still fragmented and outdated. Although some legal systems are more resilient than others to technological progress, a harmonized discipline would be beneficial to protect citizens' rights in a more efficient way, since technology often transcends the national boundaries among the Member States.

Many scholars have already advocated for a revision of the existing product liability regime due to its multiple issues, and we believe that issues regarding open source software and continuous-learning techniques should be taken into particular consideration.

The liability discipline should also be updated to include the possibility of compensation regarding all sorts of damages.

References

Alheit, K. (2001) The applicability of the EU product liability directive to software. Comparative and International Law Journal of Southern Africa 34(2):188–209

Amidei, A. (2019) Intelligenza Artificiale e "product liability": sviluppi del diritto dell'Unione Europea. In: Gabrielli E, Ruffolo U (eds) Intelligenza Artificiale e diritto, Giurisprudenza Italiana, vol 7, UTET Giuridica, pp 1715–1726

Asaro, P. (2012) On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. International Review of the Red Cross 94(886):687–709

Barbosa, M. M. (2020) O futuro da responsabilidade civil desafiada pela inteligencia artificial: as dificuldades dos modelos tradicionais e caminhos de solucao. Revista de Direito da Responsabilidade 2

Beran, T. N., Ramirez-Serrano, A., Kuzyk , R., For, M., Nugent, S., (2011) Understanding how children understand robots: Perceived animism in child–robot interaction. International Journal of Human-Computer Studies 69(7-8):539– 550

Calo, R., Froomkin A. M., Kerr I. (2016) Robot law. Edward Elgar Publishing

Char, D. S., Shah, N. H., Magnus, D. .(2018) Implementing machine learning in health care—addressing ethical challenges. The New England journal of medicine 378(11):981

Danaher, J. (2016) The threat of algocracy: Reality, resistance and accommodation. Philosophy & Technology 29(3):245–268

Expert Group on Liability and New Technologies – New Technologies Formation (2019) Liability for Artificial Intelligence and other emerging digital technologies. European Commission

Falletti, E. (2020) Decisioni automatizzate e diritto alla spiegazione: alcune riflessioni comparatistiche. Il diritto dell'informazione e dell'informatica 3:169– 206

Gabrielli, E., and Ruffolo, U. (2019) Intelligenza artificiale e diritto. Giurisprudenza Italiana 7:1657–1660

Hallevy, G. (2010) I, robot-I, criminal: When science fiction becomes reality: Legal liability of ai robots committing criminal offenses. Syracuse Journal of Science and Technology Law, 1

Hanson, F. A. (2009) Beyond the skin bag: On the moral responsibility of extended agencies. Ethics and information technology 11(1):91–99

Kamyshansky, V. P., Rudenko, E. Y., Kolominetz, E. A., Osadchenko, E. O. (2019) Regarding the issue of the essence of legal treatment and the possibility of granting legal status to a robot in civil law. In: Ubiquitous Computing and the Internet of Things: Prerequisites for the Development of ICT, Springer, pp 299–306

Karner, E. (2019) Liability for robotics: current rules, challenges, and the need for innovative concepts. In: Liability for Artificial Intelligence and the Internet of Things, Nomos Verlagsgesellschaft mbH & Co. KG, pp 117–124

Kwak, S. S., Kim, Y., Kim, E., Shin, C., Cho, K. (2013) What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In: 2013 IEEE RO-MAN, IEEE, pp 180–185

Lakkaraju, H., Kamar, E., Caruana, R., Leskovec J. (2019) Faithful and customizable explanations of black box models. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp 131–138

Marchant, G. E., and Lindor, R. A. (2012) The coming collision between autonomous vehicles and the liability system. Santa Clara Law Review 52:1321

Marchisio, E. (2020) Medical civil liability without deterrence: Preliminary remarks for future research. Journal of Civil Law Studies 13(1):4

Marchisio, E. (2021) In support of "no-fault" civil liability rules for artificial intelligence. SN Social Sciences 1(2):1–25

Matthias, A. (2004) The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and information technology 6(3):175– 183

Mi, F., Kong, L., Lin, T., Yu, K., Faltings, B. (2020) Generalized class incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp 240–241

Nyholm, S. (2018) Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci. Science and engineering ethics 24(4):1201–1219

Oh, S. J., Schiele, B., Fritz, M. (2019) Towards reverse-engineering black-box neural networks. In: Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Springer, pp 121–144

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., Wermter, S. (2019) Continual lifelong learning with neural networks: A review. Neural Networks 113:54–71

Piana, C. (2006) Licenze pubbliche di software e contratto. I contr 7:720–727

Rosenthal-von der Putten, A. M., Kramer, N. C., Hoffmann, L., Sobieraj, S., Eimler, S. C. (2013) An experimental study on emotional reactions towards a robot. International Journal of Social Robotics 5(1):17–34

Rahwan, I. (2018) Society-in-the-loop: programming the algorithmic social contract. Ethics and Information Technology 20(1):5–14

Santoro, M., Marino, D., Tamburrini, G. (2008) Learning robots interacting with humans: from epistemic risk to responsibility. Ai & Society 22(3):301–314

Scherer, M. U. (2015) Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. Harvard Journal of Law & Technology 29:353

Sharkey, N. (2010) Saying 'no!' to lethal autonomous targeting. Journal of military ethics 9(4):369–383

Sousa Antunes, H (2021) Civil liability applicable to artificial intelligence: a preliminary critique of the European Parliament resolution of 2020. Portuguese Law Review 5(1)

Tigard, D. W. (2020) There is no techno-responsibility gap. Philosophy & Technology, pp 1–19

Wolpert, T. G. (1993) Product liability and software implicated in personal injury. Defense Counsel Journal 60:519

Zipp, J. W. (2016) The road will never be the same: A reexamination of tort liability for autonomous vehicles. Transportation Law Journal 43:137