Rodríguez de las Heras Ballell, T. (2019).Legal Challenges of Artificial Intelligence: Modelling the Disruptive Features of Emerging Technologies and Assessing the Possible Legal Impact . *Uniform Law Review*, 24(2), pp. 302-314.

DOI: 10.1093/ulr/unz018

Legal Challenges of Artificial Intelligence: Modelling the Disruptive Features of Emerging Technologies and Assessing the Possible Legal Impact

Teresa Rodríguez de las Heras Ballell[*]

Abstract

The extensive use of Artificial Intelligence (AI) tools and systems and its extraordinary relevance in a multitude of social and economic domains must be framed into the broader context of a second wave of digital transformation. AI embodies the transformative force and the disruptive potential of a second generation of technologies that are ushering in a new stage of the digital evolution of our societies and economies. The acceleration and accumulation of technological developments pose unforeseen challenges to the twenty-first century's law. A systematic, extensive, and wisely combined application of these emerging technologies, such as AI and advanced robotics, Internet-of-Things (IoT), and DLT, offers fascinating possibilities and announces great disruptive effects. The aim of this paper is to devise an analytical framework to identify the disruptive features of AI, as one of the most illustrative exponent of the second-generation technologies, and assess the potential impact on certain existing principles, rules and concepts.

I.- The Meaning and the Extent of the Second Generation of Digital Transformation

The extensive use of Artificial Intelligence (AI) tools and systems and its extraordinary relevance in a multitude of social and economic domains must be framed and understood in the broader context of a second wave of digital transformation. AI embodies the transformative force and the disruptive potential of a second generation of technologies that are ushering in a new stage of the digital evolution of our societies and economies.

[*] Associate Professor of Commercial Law, Universidad Carlos III de Madrid, Spain.

In response to the first generation of digital technologies, rules on electronic commerce, digital signature and information society services were adopted[1] throughout last decades to face the legal challenges they posed to legacy concepts and principles. The technology-neutrality approach has successfully tackled the first-generation technology wave through a serene and smooth process of adjustment of legacy legal structures to the functionally-equivalent technological responses. The advent of a second wave of digital technologies seems, however, to surpass the assumptions of the existing rules devised to be accommodated to the characteristics of first-generation technologies. The second technological generation challenges the well-settled principles of technological neutrality and functional equivalence[2] that are the backbone of the current international legal framework for the digital economy.

The acceleration and accumulation of technological developments pose unforeseen challenges to the twenty-first century's law. A systematic, extensive, and wisely combined application of these emerging technologies, such as AI and advanced robotics, Internet-of-Things (IoT), and DLT, offers fascinating possibilities and announces great disruptive effects. A technology-neutral approach might not be enough, not even necessarily advisable, to embrace this second wave of technological innovation.[3]

Therefore, the hypothesis of this paper is that legal and practical challenges may arise from a set of disruptive features that emerging technologies present where compared to previous technological advances. Certainly, although such distinctive features of emerging technologies are not entirely new in relation to other technological developments in the past, there are two factors that reinforce their disruptive character. First, many of the disruptive effects of today's new technologies emerge or, at least, are

[1] UNCITRAL *Model Law on Electronic Commerce* 1996 with Guide to Enactment (MLEC), UNCITRAL *Model Law on Electronic Signatures* 2001 (MLES), *United Nations Convention on the Use of Electronic Communications in International Commerce* (CEC) approved by General Assembly Resolution 60/21, of 23 November 2005 (www.uncitral.org). Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce).

[2] Principles enshrined in UNCITRAL MLEC, UNCITRAL MLES, UNCITRAL CEC. These principles have guided the subsequent instruments adopted by UNCITRAL. An illustrative example is the recent UNCITRAL Model Law on Electronic Transferable Records (2017).

[3] Assessing that inadequacy of a technology-neutral approach in relation to secured transactions laws, Rodríguez de las Heras Ballell, Teresa, "Digital Technology-Based Solutions for Enhanced Effectiveness of Secured Transactions Law: The Road to Perfection?", *Law and Contemporary Problems, Duke University School of Law,* Vol. 81, num. 1, 2018, pp. 21-44.

exacerbated by the combination of several technologies in a complex ecosystem – interconnecting IoT-driven devices, DLT, Artificial Intelligence applications, algorithm-driven smart contracts, etc. Second, existing features become highly disruptive as they reach a 'point of disruption' that constitutes a point of inflexion in the previous analysis. That means that although certain features of emerging technologies are not unique or singular, they manifest today in a higher degree, at a larger scale, and, therefore, existing concepts, rules and principles might be not fully suited to embrace them. As it will be further explained below, the level of autonomy, the degree of complexity, or the increasing vulnerability deriving from data dependency or exposure to cyber risks in the ecosystem of emerging technologies turn into new disruptive features that need special attention and could require clarifications, adjustments or even reconsideration of certain traditional legal concepts, principles and rules.

Thus, the legal implications of the use of AI-driven systems has to be framed in a wider reflection on the legal challenges that the second-generation technologies give rise to. This proposal of contextualization of the analysis entails that new rules should not be formulated simply to respond to the emergence of a specific technology. To the maximum extent possible, the advent of a new technological development should be embraced, at a first stage, by a technology-neutral approach and a functional-equivalence logic. Only where a technological advance proves to be disruptive enough, then a testing phase of existing concepts and rules have to be opened. I sustain that the disruptive potential of technologies derives from certain substantive, operational or structural features that prevent from simply extending to them the functional-equivalence solution, as it has very effectively worked to embrace technological advancements to date. In those cases, efforts to fit disruptive technologies into existing technology-neutral concepts and to extend the application of functional-equivalence rules might fail.

Such a perspective guarantees that the legislative reaction to the development of new technologies does not produce technology-specific rules at the pace set by the market, the technological progress, and the business practices. Those responses would become obsolete upon the arrival of a substituting competitive technology or the simple loss of popularity of a specific technological solution. On the contrary, steps towards a legislative action on disruptive technologies should be mindful and deliberate to identify disruptive features of arriving technologies and unveil their legal implications. To my

mind, a hurried action should be very much discouraged, as it can disrupt the market and alter the adoption process of emerging technologies by business actors. Furthermore, I anticipate here one of my conclusions. Should a legislative action be deemed necessary, a harmonized response, preferably at an international level, is highly advisable. In absence of harmonization, international trade incorporating emerging technologies would be hindered, technological progress and penetration could be distorted, and jurisdictional arbitrage would be fostered. Uniform principles, or rules, where appropriate, are highly desirable.

The aim of this paper is to devise an analytical framework to identify the disruptive features of AI, as one of the most illustrative exponent of the second-generation technologies, and assess the potential impact on certain existing principles, rules and concepts. To that end, the paper is structured in three parts, in addition to this introductory section (Part I). In Part II, the key distinctive features of second-generation technologies are proposed. The main In Part III, considering the distinctive features that have been previously outlined, it is assessed whether existing concepts and rules can still be applied or a process of reconceptualization is needed in the realm of Contract law and Civil Liability. Part III concludes by summarizing final remarks that revolves a fundamental conclusion: the need for harmonization in the formulation of rules for disruptive technologies.

II.- The disruptive potential of Artificial Intelligence: Risks, Opportunities and key Disruptive Features

In this paper, the term Artificial Intelligence (AI)[4] systems refers to systems, based on algorithms and self-learning guided by Machine Learning and Deep Learning, able to perform certain human cognitive capabilities by interacting with the environment through

[4] The term of AI was firs coined in 1955 by John McCarthy who organized the famous Darthmouth Conference (*Dartmouth Summer Research Project on Artificial Intelligence*) in 1956. McCarthy, J., "Programs with Common Sense", in *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, London: Her Majesty's Stationery Office, 1959, pp. 756-791. The famous paper published by Alan Turing, "Computing Machinery and Intelligence", *Mind,* num. 49, 1950, pp. 433-460, is considered a key milestone in the development of researches and publications in the academic field.

sensors, processing information, and adopting decisions and taking actions, with a certain (increasing) degree of autonomy.[5]

The definition of AI immediately leads to the concept of algorithm. An algorithm is a finite sequence of instructions, rules or actions to solve a problem. Hence, an algorithm-driven system constitutes a structured process to provide a solution to any instance of a recurrent problem. In that regard, the implementation of algorithm-driven systems enables to automate a variety of decisions, tasks, or processes – classifying, searching, scoring, ordering, ranking, selecting, filtering. Possible applications are multifarious from financial services[6] (high-frequency trading, robo-advisers or automated financial advice, algorithmic trading)[7] to other uses in an immense array of non-financial sectors (chatbots, personalized advertisements / contextualized marketing, recommender systems, customer relations, university rankings, complaint handling, medical diagnosis, dispute resolution, recruitment, etc).[8]

A basic concept of algorithm encapsulates the idea of programming through instructions. Instruction-based programming entails predictability, as the outcomes are essentially the expected results of the pre-conditions, the decision criteria and the algorithm design. As MacCarthy explained "(a) machine is instructed mainly in the form of a sequence of imperative sentences".[9] Nevertheless, AI does require a shift from actions based on instructions to decisions based on rules. Machine Learning and Deep Learning guide the learning of an algorithm-driven system. Machine Learning is essentially based on the parsing of data to learn, predict, and adopt a decision on the basis of a set of variables. Deep Learning is a technique within Machine Learning tools that

[5] "Strong AI" describes the envisioned development of machines and autonomous systems able to perform cognitive capabilities and intellectual abilities equivalent to and indistinguishable from human beings. Current state-of-the-art technological progress has proved to produce to date machines and smart systems based on algorithms and guided by machine learning and deep learning. That is defined as "Weak AI", "Applied AI" or "Narrow AI".

[6] Flynt, Oscar, *Fintech: Understanding Financial Technology and Its Radical Disruption Of Modern Finance,* 2016

[7] PwC, *Blurred lines: How Fintech is shaping Financial Services, Global FinTech Report,* 2016. International Monetary Fund, *Fintech and Financial Services: Initial Considerations,* IMF Staff Discussion Note, June 2017, SDN/17/05, World Economic Forum, *Beyond Fintech: A Pragmatic Assessment Of Disruptive Potential In Financial Services* (Aug. 22, 2017) available at https://www.weforum.org/reports/beyond-fintech-a-pragmatic-assessment-of-disruptive-potential-in-financial-services (last visited 17/09/2018).

[8] Scholz, Lauren Henry, "Algorithmic Contracts", *Stanford Technology Law Review,* num. 20, 2017, pp. 101-139

[9] McCarthy, J., "Programs with Common Sense", *op.cit.,* p. 4.

aims to enable example-based learning of machines and autonomous systems. Instead of instructing the system with a set of pre-determined instructions, Deep Learning provides a model for the machine to evaluate examples and infer patterns for the solving of future problems.

Benefits from the use of algorithms and AI are numerous.[10] Algorithm-driven systems provide with celerity, simplicity, and effectiveness the solving of a multitude of problems. Yet, automation reduces transaction costs dramatically enabling the provision of services in reasonable conditions that were unprofitable, unaffordable, or unfeasible in other circumstances. Cost-reduction factors[11] explains, for instance, the burgeoning sector of robo advisers that have expanded the market beyond the traditional financial advisers benefitting consumers, diversifying the offer, increasing competition, and enhancing financial inclusiveness.[12] Such an expansion has facilitated the delivery of financial advice to small investment patrimonies, and lower-income investors in market conditions. Therefore, algorithm-driven systems can perform automated tasks and adopt mass decisions in an efficient way (High Frequency Trading, search engines, face recognition, personal assistant, machine translation, predictive algorithms, recommender systems). The use of algorithms is critical for the provision of a number of key services in our society on a mass scale that would be impossible or highly inefficient otherwise (searching, classifying, filtering, rating, ranking).[13]

Nonetheless, the use of algorithms might also give rise to undesired or unexpected consequences, pose risks and can arouse societal concerns and legal challenges.

[10] Deloitte, *Artificial Intelligence. Innovation Report 2018.*

[11] Robo-advisers are perceived as a low-cost alternative to financial advisors. In its report, Deutsche Bank Research, 'Robo-advice – a true innovation in asset management, *EU Monitor, Global financial markets,* August 10, 2017, p. 9, comparative data on average fees applied by traditional financial advisors and robor-advisers both in the US and in Europe are provide validating that perception.

[12] Although affordability and accessibility to financial services (i.e. low-cost automated portfolio management, robo-advisers) are significantly enhanced by digital financial innovation, financial inclusiveness depends upon other factors more related to financial literacy or wealth. Bucher-Koenen, Tabea and Ziegelmeyer, Michael Heinrich, Who Lost the Most? Financial Literacy, Cognitive Abilities, and the Financial Crisis (January 10, 2011). MEA Discussion Paper No. 234-11. Available at SSRN: https://ssrn.com/abstract=1738368 or http://dx.doi.org/10.2139/ssrn.1738368; Bucher-Koenen, Tabea, *Financial Literacy, Cognitive Abilites, and Long-term Decision Making. Five Essays on Individual Behavior*, Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim, 2010

[13] COM(2018) 795 final, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Coordinated Plan on Artificial Intelligence. Brussels, 7.12.2018.

Algorithmic decisions can be biased or discriminatory[14] as a result of prejudicial pre-conditions, inadequate design, or insufficient or ill-selected set of data.[15] As algorithm-driven systems adopt automated decisions, the possible bias becomes massive and, in many cases, is unnoticed while goes viral. In a densely-interconnected society, network-based virality act as an amplifier of harmful effects. Negative impact expands quickly, the magnitude of the damage increases vastly, and the reversibility of the actions become more unlikely and less feasible.

Furthermore, sophisticated AI-assisted systems add additional uncertainty, as the learning process generates increasing unpredictability to the system's responses to future situations. As higher and broader the learning capabilities are, more unpredictable the outcomes could become. The synergic combination of self-learning, increasingly autonomous decisions, dependency on data, and growing complexity obscure the allocation of risks and liability on the basis of traditional principles

Negative effects of algorithms and AI can have multiple manifestations. AI powered systems in varied sectors can cause significant patrimonial damages and personal injuries– i.e. autonomous vehicles, remotely piloted drones, smart homes, healthcare robotics. Algorithm-driven financial services can trigger systemic risks, undermine the financial market stability, or generate market shocks with unpredictable consequences. While the deployment of algorithms in rankings, recruitment services, digital-content filtering or chatbot for complaint handling can have severe impact on free speech, right to equality and non-discrimination, honor and reputation, personality rights or market competition.

In the face of such potential negative impact, it has to be assessed whether legacy legal regimes are poised to manage the risks, and effectively solve the conflicts arising from all those situations. To that end, common distinctive features[16] must be identified and compared with previous situations to which existing rules are well accommodated.

[14] Chander, Anupam, "The Racist Algorithm, 115 *Michigan Law Review*, 2017, pp. 1023-1046.
[15] Barocas, Solon & Selbst, Andrew D., "Big Data's Disparate Impact", 104 *California Law Review,* 2016, pp. 671-732
[16] Commission Staff Working Document *Liability for emerging digital technologies.* Accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions *Artificial intelligence for Europe* {COM (2018) 237 final}, Brussels, 25.4.2018 SWD(2018) 137 final, Annex I.

Such distinctive features would allow to model a category of disruptive technologies where to test the adaptability and suitability of legacy legal rules. Against such a backdrop, it is important to note that the analytical model based on a set of disruptive features, that it is proposed here, does solely aim to systematize the disrupting potential of technology, but the acknowledgement of the disruptive character is not correlated necessarily with the need for new rules. A legitimate policy decision could be to subject disruptive technologies to the same rules as preceding technologies on the grounds that technological neutrality is a preferred principle.

The following four distinctive features encapsulate the disruptive potential of AI, especially in combination with other converging technologies.

1). Complexity.

Emerging technologies, specially integrated in sophisticated technological ecosystems, show a considerable level of complexity. Such a complexity manifests in three layers: internal logical complexity, plurality of participants and sources contributing to the operation of the system, and ecosystem of connected objects (sensors, actuators, networks, softwares, oracles, data collectors, platforms).

Algorithms driving sophisticated autonomous systems imply a high level of complexity in the design as well as in the operation. That adds opacity to the internal processing of the autonomous system, conceals the relevant criteria for the decision making, and reduces the comprehensibility of the outcomes. Yet, the opacity of the algorithm/AI schemes, due to the complexity and the lack of transparency of the whole procedure, normally entails the very unawareness by the addressee of the pre-conditions, the criteria, and the procedural aspects of the algorithmic decision.

Complexity does also manifest externally. In the design, the operation, and the functioning of these ecosystems, a plurality of actors can participate or be anyhow involved: software and app developers, algorithms' designers, data providers, sensors manufacturers, system operators, producers of each device, part or component, DLT providers, monitoring service providers. Besides, complexity does also describe the multiplicity of parts, components, devices and systems integrating a technological

ecosystem – i.e., an autonomous car, a sophisticated surgical robot, a connected smart home system, or an algorithm-driven automated financial advisor.

2). Increasing Autonomy.

The second challenge is linked to the level of autonomy and the machine-learning capabilities[17] that algorithm-driven systems, as intelligent agents, may have. Increasing autonomy of algorithm-driven[18] systems constitutes one of the most disruptive factors of the second-generation technologies. Autonomy[19] is, nevertheless, a degree of a scale. It must be defined at which point traditional solutions for the allocation of legal effects and the attribution of liability[20] become inadequate[21] and new solutions are needed.

Contrarily, if this caracteristic feauture of AI is deemed, despite the novelty and apparent intensity, purely incremental, an effort to preserve the current system, under the guidelines of technological neutrality and functional equivalence, should be the expected and desired response. In that sense, the disruptive features highlighted in this Paper are aimed to describe the challenges that the convergence of these emergence technologies poses, but, as it was anticipated, they are not necessarily determining a disruptive legal response to such challenges.

3). Opacity

Increasingly complex algorithms drive autonomous systems with self-learning capabilities that select candidates for a job, a loan, or a grant, build consumer profiles, classify and filter digital content, group users, or redirect spam, fake news, or

[17] Burrell, J. (2016) "How the Machine 'Thinks:' understanding opacity in machine learning algorithms", *Big Data & Society,*2016, pp. 1-12.
[18] Scholz, Lauren Henry, "Algorithmic Contracts", *Stanford Technology Law Review,* num. 20, 2017, pp. 101-139.
[19] European Parliament, *European Civil Law Rules in Robotics, Study for the JURI Committee,* PE 571.379, 2016.
[20] Kroll, Joshua A., *et alii,* "Accountable Algorithms", *University of Pennsylvania Law Review*, num. 165, 2017, pp. 634-706.
[21] Rodríguez de las Heras Ballell, Teresa, "Intermediación en la Red y responsabilidad civil. Sobre la aplicación de las reglas generales de la responsabilidad a las actividades de intermediación en la Red", *Revista Española de Seguros,* núm. 142, 2010, pp. 217-259; and "La responsabilidad de los prestadores de servicios de intermediación y los estratos de la intermediación en la Red", *Revista Derecho y Tecnología,* núm. 11, 2010, pp. 69-96

advertisement, fed by immense amounts of data collected by a variety of sources and inferred from behaviours, social interaction with presumed like-minded circles, past experiences[22] and transactions.[23] Criteria basing the decisions are often unknown and the design of the underlying process opaque.[24] Lack of transparency exacerbates the complexity and the uncertainty to allocate liability.

Sophisticated algorithm-driven systems operating a technological ecosystem – an autonomous car, a smart home system, or a robo adviser in financial markets - are not transparent ('black box effect').[25] The complex set of instructions, criteria, weight factors, data or alternative options is not normally visible (nor easily understandable) for the end user. But, more importantly, in many cases, the mere transparency of such elements would not ensure sufficient comprehension of the criteria leading the decision-making, the reasons of malfunctioning, or the causes provoking the damage. In sum, the explainability[26] of complex technological systems is limited, costly, and not always fully feasible in the whole extent.

The idea of transparency casts over several aspects of an algorithmic decision.[27] On the one hand, transparency could be referred to the accessibility and the comprehensibility of the design and the operation of the algorithm-driven process, as well as the criteria or the reasons taken into account to produce the adopted decision. In that regard, the reference to explainability could be related to the general design of the algorithm in abstract terms or to every single decision adopted thereby. On the hand, transparency could also comprise the question whether the fact that a task is performed

[22] Dwork, Cynthia et alii, "Fairness Through Awareness", 2012 *Proceedings 3rd Innovations Theoretical Computer Science Conference*, 2012, pp. 214-226

[23] The opacity is not only attributable to the complexity of the algorithm processing the information and producing outcomes based thereon, but also to the datasets fueling the decision-making. The diversity and plurality of data gathered and process by Big data analytics tools hamper the explainability, accountability, and comprehensibility of the algorithm-based decisions - *Joint Committee Discussion Paper on the Use of Big Data by Financial Institutions*, produced by EBA, EIOPA, and ESMA (the ESAs), JC 2016 86, p. 27.

[24] Ananny, Mike, "Toward an Ethics of Algorithms Convening, Observation, Probability, and Timeliness", *Science, Technology & Human Values*, vol. 41, 2016, 1-25.

[25] Pasquale, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge-London, 2015.

[26] Goodman, Bryce and Flaxman, Seth, "EU Regulations on Algorithmic Decision-Making and a "Right to Explanation"', arXiv:1606.08813 (https://arxiv.org/pdf/1606.08813v2.pdf, last visit 12/3/2019), 2016.

[27] Wachter, Sandra and Mittelstadt, Brent and Floridi, Luciano, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation" (December 28, 2016). *International Data Privacy Law,* 2017. Available at SSRN: https://ssrn.com/abstract=2903469 or http://dx.doi.org/10.2139/ssrn.2903469.

or a decision is adopted by an automated/autonomous system, instead of a human being, should be disclosed, in any case or upon the request of the addressee. That approach lead to the debate about a right to be informed or a right to receive an explanation in general terms or in decision-specific ones instead.

4). Vulnerability

AI systems are technologically vulnerable. Vulnerability refers to two situations. On the one hand, AI systems are amply dependent upon data – collected data, test data, learning data, processed data, machine-generated data, user's data, personalizing data. Data determine the accuracy of the outcomes, fuel the decisions, feed the machine-learning process, and ensure the very operation of the system. Data dependency is a source of vulnerability. Insufficient, inaccurate, or biased data compromise the performance of the AI system. On the other hand, AI systems are exposed to cybersecurity attacks or breaches. In sophisticated AI systems driving complex technological ecosystems – autonomous drones, autonomous vehicles, smart home systems –, the consequences of a cybersecurity breach can be immense.

All these distinctive features are increasingly disruptive. That dissuades from simplifying the analysis of new technologies. They are not simple incremental evolution of previous technologies. In some aspects, they reach the 'point of disruption', what invite clarifications, adjustment or reconsideration of existing concepts, rules, and methods.

Given the wide use and the variety of applications of AI, relevant disruptive impact can extend over a multitude of rules and regulations, such as privacy, competition, financial regulation, safety standards, defective product, fundamental rights, contract law, civil liability. Among them, the subsequent section exclusively addresses a few legal issues related to contract law and liability. Those areas could be the most relevant, although not the only ones, in the devising of a project for the international unification of private law related to AI.

III.- Possible Legal Areas of Impact and Legal Responses: A Case of Harmonization.

The above described features depict the disruptive potential of AI in a variety of legal aspects. A thorough analysis of all these possible impact points is beyond the aim of this Paper and is indeed work in progress in multiple legal fields. Hence, the sole purpose of this final part of the Paper is to sketch an impact assessment by simply delimiting and grouping points of frictions triggered by the disruptive features with no further detail or specification. Finally, possible legal responses are also briefly exposed on the basis of the taxonomy of challenges.

Such an exercise would at least reveal and spotlight which concepts, legal schemes, or legacy structures are more likely to be challenged. Clarifications, adjustments or adoption of new rules could follow from that revelation.

First, disruptive features of AI, as they have been outlined above, have an evident impact of tort law. Some key concepts underpinning traditional liability regimes could be anyhow shaken by such disruptive features. That might render existing regimes insufficient or partially inadequate. The adequacy and completeness of liability regimes in the face of technological challenges have an extraordinary societal relevance. Should the liability system reveal insufficiencies, flaws and gaps in dealing with damages caused by emerging technologies, victims may end up totally or at least partially uncompensated. The social impact of a potential inadequacy of existing legal regimes to address new risks created by AI might then compromise the expected benefits and aggravate the social perception of risk undermining the acceptance rate of emerging technologies.

From the combination of complexity and opacity, practical problems and legal challenges immediately arise.

In all its facets, the increasingly high complexity embedded in new technologies' applications triggers an evident practical problem with legal relevance. Multiple actors could contribute to the causation of the damage. A plurality of actual or potential tortfeasors is certainly not a new problem for tort law that provides indeed for solutions.[28]

[28] Koch, Bernhard A., "Proportional Liability for Causal Uncertainty", in Martin-Casals/ Papayannis, *Uncertain Causation in Tort Law*, Cambridge University Press, 2015; Magnus, U., "Multiple Tortfeasors under German Law", in Rodgers (ed.), *Unification of Tort Law*, The Hague, 2004; Winiger/Koziol/Zimmermann/Koch (eds.), *Digest of European Tort law I: Essential Cases on Natural*

However, in these cases, the multitude of players could act without prior coordination or planned intervention, the contribution could be occasional or spontaneous, and the participation of some players can be totally unknown or even unforeseen by the main operators (data provider, hacker, non-interoperable system, unexperienced user). Hence, in some circumstances, flawed functioning, harmful outcomes, or damaging operation of the system can be provoked by lack of interoperability among components, interaction with other unexpected components or software, wrong data, or inexpert or inadequate use by the user. In such scenarios, the damage cannot be easily attributed to a specific component, a well-defined cause, or a single actor. The damage derives from the ecosystem as a whole.

Far from the classical monocausal conception of causing harm, the increasingly complex AI system reveals also a plurality of possible causes. Frequently, the damage results from a conjunction of interweaved effective causes and has been collectively triggered by multiple actors. This situation is not unfamiliar to current legal systems either. Rules dealing with damages caused by multiple causes are indeed presently provided for in all jurisdictions. Nonetheless, AI systems add further intricacies to that well-known problem. Unlike traditional products, once circulated by the original manufacturer and without the latter's participation, subsequent actors can intervene in the marketing, the use, or the upkeep of sophisticated technological products without participation of the original manufacturer. Accordingly, in the cycle of production-use of the product subsequent activities, tasks, and causes can interfere and contribute in the likeliness of the causing of damage. That multi-layer process implies the overlapping and convergence of many sources of damage – software updates, personalizing options by the end user, self-learning actions, data collection.

Opacity adds further complexity thereto. In a context of low transparency and limited explainability, it is difficult to unveil the cause. Not surprisingly, the process of discovery and evidence becomes costly and complicated, and not always feasible.

The vulnerability feature signals other weak points of AI systems and, therefore, the magnitude of the exposure. Dramatic personal injury can be caused by a poorly-

Causation, Vienna, 2007; Winiger, B., "Multiple Tortfeasors", in Tichý (ed.), *Causation in Law*, Praha, 2007.

performing surgical robot due to wrong data or a hacking attack. Likewise, the consequences of a cybersecurity breach disrupting the operation of a fleet of autonomous drones or autonomous vehicles can be catastrophic. Furthermore, liability impact could be also aggravated by the multiplying effect of automation and virality. The magnitude of the harm caused by AI magnifies, whereas damages can easily become viral and rapidly propagate in a densely-interconnected society.

Finally, autonomy captures one of the most perturbing effect on the classical discourse articulating liability regimes. The classical fault-based liability rules are inspired by an anthropocentric conception. Concepts such as fault, conduct, intention or standard of care have been conceived, developed, applied, and interpreted essentially for and in relation to humans. Whom to attribute liability to if a harmful outcome is not predetermined by the programming but the result of an "autonomous decision" of the AI system? How to apply the notions of fault to the "conduct" of autonomous systems? What is the standard of care to assess the operation of an AI-driven autonomous system? None of these questions are unanswerable. Even more, the answers could not be necessarily disruptive.[29] A continuity approach is a valid and legitimate option. But a debate is needed. Consequences of alternative policy decisions should be considered and duly pondered – preservation of current liability regimes, orientation towards strict liability models, mandatory insurance, extension of defective product liability regimes, formulation of standards, creation of sectoral compensation funds, legal recognition of electronic personhood. Effects on innovation, production costs, acceptance rate of emerging technologies by population, and robustness of the system have to be assessed and included in the policy decisions equation.

Second, increasing autonomy of AI powered systems, fueled by expanding self-learning capabilities, raises appealing questions related to the classical understanding of the rules for commercial transactions.[30] Both the possibilities and the limitations of self-executing "smart contracts" invite to reflect on many basic foundations of contract law. Throughout the stages of the contract formation process, interesting questions arise. From

[29] European Parliament, Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, P8_TA-PROV(2017)0051, para 59.
[30] A comprehensive study of contractual issues related to smart contracts in Feliu Rey, Jorge, Smart Contract: Concepto, ecosistema y principales cuestiones de Derecho privado", *La Ley mercantil,* num. 47, 2018, pp.1-25.

comprehensibility of the machine language, concordance of the coding with the parties' actual intent, and issues related to mistake or lack of consent, to incorporation of standard terms, interpretation and gap-filling, or limitations in the use of general principles and standards (standard of care, reasonableness, good faith, best efforts), and to the limits for the execution of self-help measures and other automatic technology-enabled remedies in case of default. Some concerns can be addressed and solved from a functional-equivalence approach, others may require more groundbreaking solutions.

Third, the increasing automatization of a multitude of activities, decision-making and tasks with legal relevance and actual impact on individuals' rights stirs a profound debate about the full exercise of our private autonomy in a world of expanding autonomous systems – personalized advertisement, preselected news, 'ideological silos', available services or pricing options on the basis of user profile, profiling. In the face of those challenges to private autonomy and other rights and liberties, it might be argued that the configuration of new rights, either as an evolution of existing rights to be accommodated to the new environment or as genuine new rights,[31] is needed.[32]

The ideas outlined above are purely exemplification of possible areas of friction that might require legislative attention: a profound rethinking of the very concept of contract and its role in society, a recalibration of remedies, the setting of standards for AI-based decisions to be valid and enforceable, the refashioning of dispute resolution mechanisms, and the incorporation of technologies in prevention and civil enforcement.[33] There is no aspiration other than point out some areas of attention and evaluate the utility of the model of disruptive features for emerging technologies in the test process. A deep, meticulous, and detailed analysis is absolutely instrumental to a proper assessment of the adequacy of our existing legal regimes.

[31] Among other proposals, a general right to access to digital services is discussed, as well as a set of rights related to automated decision-making based on profiling under the European General Regulation on Data Protection (Article 22) – right to explanation, right to object, right to human intervention.

[32] In the joint publication, De la Quadra Salcedo (Dir), *Sociedad Digital y Derecho,* Madrid, BOE, 2018, the authors address from an multidisciplinary perspective the diversity of challenges that the ongoing building of a digital society must face and discuss the need for the enshrining of new rights to effectively protect individuals' interests.

[33] As I have previously advocated for in Rodríguez de las Heras Ballell, Teresa, "Digital Technology-Based Solutions for Enhanced Effectiveness of Secured Transactions Law: The Road to Perfection?", *Law and Contemporary Problems, Duke University School of Law,* Vol. 81, num. 1, 2018, pp. 21-44, at p.44.

Upon defining the contours of possible areas of impact of AI, a roadmap for policymakers with alternative legal responses can be traced. Essentially, the first policy decision should address the adequacy of a technological-neutrality approach to face the disruptive features of AI. Should extending the technological neutrality principle to AI be deemed the preferred response, efforts had to focus on providing clarifications to ensure the applicability of existing concepts and rules to AI scenarios. Maintaining the functional-equivalence technique to that end would seem a very reasonable option. On the contrary, whether, due to the disruptive character, legal challenges posed by AI cannot be contained within the perimeters of the technological-neutrality and functional-equivalence principles, other legal approaches must be explored. In that case, a strong case for harmonization is made to prevent regulatory fragmentation that might undermine innovation, promote jurisdictional arbitrage, enable undesired applications of AI, and hamper international trade. First, the formulation of uniform principles for the use of emerging technologies in international commercial contracts. Second, a model law on a selected of legal issues (smart-contract-related contractual issues, technology-enabled automatic remedies, liability rules for AI-driven systems) to promote harmonized domestic responses. Third, a binding instrument aimed either to supplement existing uniform texts where AI applications are used to transnational transactions, or to provide a complete regime for AI-driven systems in an international context covering a limited array of relevant issues (contractual, liability, dispute resolution, property issues).