Preface

As digitalisation has become a fundamental trend changing our economy into a digital economy, EU legislation is increasingly faced with the task to provide for a legal framework, a European Digital Single Market allowing to reap economic growth from digitalisation. While attention so far has mainly been paid to contract law, challenges obviously extend beyond this area of law. This becomes particularly clear with a view to Artificial Intelligence (AI). Being a key driver in building a digital economy, AI not only is an important factor for reaping economic growth but also brings about risks that have to be dealt with.

In accordance with the aim of the "Münster Colloquia on EU Law and the Digital Economy" to discuss how EU law should react to the challenges and needs of the digital economy, the 4th Münster Colloquium, held on 12–13 April 2018, focused on possible EU law responses to such risks arising from the use of AI. With "Liability for Robotics and in the Internet of Things" the Colloquium not only addressed questions relating to the reasonable allocation of these risks but also shed light on possible forms of liability taking into account traditional concepts of liability as well as possible new approaches.

This volume collects the contributions to this fourth Münster Colloquium. The editors kindly thank Karen Schulenberg for her invaluable support in organizing the Colloquium and in preparing this volume.

October 2018

Sebastian Lohsse
Reiner Schulze
Dirk Staudenmayer

Contents

Introduction

Multilayered (Accountable) Liability for Artificial Intelligence

Giovanni Comandé[*]

I. The networking society and the internet of humans: an introduction

Today's technologies enable unprecedented exploitation of information, be it small or big data, for any thinkable purpose ensuing thus juridical and ethical anxieties. Algorithms are regularly used for mining data, offering unexplored patterns and deep non-causal analyses to those able to exploit these advances. Yet, these innovations need to be properly framed in the existing legal framework, fit in to the existing set of constitutional guarantees of fundamental rights and freedoms, and coherently related to existing policies in order to enable our societies to reap the richness of big and open data while equally empowering all players.

By referring to algorithms as a pivotal element of 'new technologies', we hereby summarize for the sake of brevity reference to the use of the so-called machine learning to produce: 1) new unexpected solutions; 2) the ability to exploit the interconnectedness of things and of humans; 3) the embedding of Artificial Intelligence – AI (autonomous more than automatic) based decision making in robots or the AI deployment as a software; and, 4) algorithms' implications in the use of 3D printing[1].

In short, today we live in a continuously expanding explosion of data production and data use. Our life is permanently connected to internet both via the direct access we have to it or by the use we make of connected objects and tools in our daily activities. Both our connections avenues

---

* Full Professor of Private Comparative Law, Scuola Superiore Sant'Anna, Pisa.

1 Anton Vedder and Laurens Naudts argue that 'Algorithmic accountability does not only require the examination of algorithms or the code as such but also an examination of how algorithms are deployed within different areas and what the tasks are that they perform' and that 'The potential interconnectedness of algorithms and of algorithmic decisions also seriously restricts the means of algorithmic decision makers to give an account of the decisions they make'. Anton Vedder and Laurens Naudts, 'Accountability for the use of algorithms in a big data environment' (2017) 31 International Review of Law, Computers & Technology 206, 209.

to internet use data (personal and non-personal ones) and generate new data, which in turn a number of entities analyse and often monetise.

Note from the outset that almost any liability regime that can be envisaged for these technologies would require for (technological) reasons the 'help' of the very same technologies it is regulating to make effective its legal rules. Thus, the emerging regulatory approach must necessarily blend various legal, technological, and economic strategies for which the time frame is of crucial importance.

Algorithms, big data, and the large computing ability that connect them do not necessarily follow causal patterns and are even able to identify information unknown to the human individual they refer to. These elements deeply alter the notion of causality employable by any chosen liability regime[2].

Datasets are actively constructed and data do not necessarily come from the same sources, increasing the risks related to de-contextualization. Moreover, since an analyst must define the parameters of the search (and of the dataset to be searched), human biases can be 'built in' to the analytical tools (even unwillingly) with the additional effect of their progressive replication and expansion once machine learning is applied. Indeed, if the machine learning process is classified as 'non-interpretable' (by humans), for instance because the machine learned scheme is assessing thousands of variables, there will not be human intervention or a meaningful explanation of why a specific outcome is reached.

In the case of interpretable processes (translatable in human understandable language), a layer of human intervention is possible (although not necessary). Again, this possibility is a double-edged sword since human intervention can either correct biases or insert new ones by interfering with the code, setting aside or inserting factors. In any event, the technological features will necessarily affect the selected liability rules in various ways.

The lack of (legal and ethical) protocols to drive human action in designing and revising algorithms clearly calls for their creation, but requires a common setting of values and rules, since algorithms are basically enjoying a-territoriality in the sense that they are not necessarily used in one given physical jurisdiction. Moreover, the expansion of the autonomous

2 Giovanni Comandé, 'The Rotting Meat Error: From Galileo to Aristotle in Data Mining? (2018) 4 European Data Protection Law Review 270–277.

generation of algorithms calls for building similar legal and ethical protocols to drive the machine generation of models. In both cases, there is also a need for technological verifiability of the effectiveness of the legal and ethical protocols making human readable at least the results of the application of the model when the model itself is not readable by humans.

Data and the ability of making use of them is the key for both developing new informatics tools, such as analytics and predictive coding, and feeding the creation of artificial intelligence (AI) and its actual operation.

As in the relationship between the human body and its brain, the brain plays the most relevant part. Similarly, in the emerging use of robotics the kingmaker is not the robot itself, but the AI enabling its performance. In a word, disruption does not come from Robotics or IoT as such. It comes from the embedding and expansion of Artificial Intelligence in products. Robots as such do not differ much from things, employees, and collaborators. This is why, until AI moved out from science fiction to appear in the real world, liability of robots themselves was not on the table.[3] However, there are several semantic misunderstandings both in the use of the word 'intelligence' and in several connected elements that need clarification before tackling the issue of liability more in detail.

What do we mean by Artificial Intelligence? AI is intelligence exhibited by machines. In computer science, the field of AI research defines itself as the study of 'intelligent agents': any device that perceives its environment and takes actions that maximize its chance of success at some goal. This definition already illustrates that artificial intelligence is mostly vertical: it is excellent in doing one thing as compared to the versatility of the human intelligence that we could call horizontal and apt to a manifold of tasks. The AI animating our cleaning robot can enable it to clean better and longer than a human can but cannot perform at all any other 'intelligent' task the worst human cleaner can perform sufficiently well.

AI largely builds on Machine Learning: the subfield of computer science that, according to Arthur Samuel, gives computers the ability to learn

---

3 For a literature review on the attribution of legal personhood and liability to non-human entities operating at an increasing distance from the physical persons, such as pseudonyms, avatars, and software agents, see Bert-Jaap Koops, Mireille Hildebrandt and David Olivier Jaquet-Chiffelle, 'Bridging the Accountability Gap: Rights for New Entities in the Information Society' (2010) 11 Minnesota Journal of Law, Science & Technology 497.

without being explicitly programmed.[4] It is Big Data to make possible Machine Learning. Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Big data are continuously generated by sensors, humans, algorithms, computers, things, etc, and increase unceasingly.

In turn, Big Data enables AI to develop degrees of 'Autonomy' (mostly vertical as the described cleaning robot): autonomy is not of a moral character but rather has to be expressed in terms of relative absence of human control. It offers unpredictable or at least unpredicted behaviors that might entail a reduction or shifting of liability from the moral (human) agent involved in the production, deployment, use of the system using the 'autonomous' AI.[5]

II. Automation, autonomy and unpredictable behaviours

This further step requires a clarification as well. We need to distinguish between 'automation' that has been with us since a very long time (think of any first industrial revolution factory, for instance) and 'autonomy' of things that is the real novelty in terms of triggering new issues in liability (think of a fully autonomous vehicle as a well-known example).

The borderlines between automated (Airports' transfer trains, for instance) and autonomous systems (fully autonomous vehicles) are progressively blurring for several reasons we cannot discuss here. Suffice is to say that autonomy depends on the margins of 'liberty' AI has in its choices. Their degrees of liberty are often guided by the information they can gather and process in their own automated decision-making processes; similarly to what humans do but at a much wider level and at higher computational speed.

Equally to humans, AI systems can operate in unstructured environments, making them dependent on sensor data and information or on the

4 Arthur L Samuel, 'Some Studies in Machine Learning Using the Game of Checkers' (1959) 3 IBM Journal of Research and Development 210.

5 Yet the limitation of criminal liability, for instance of self-driving cars operators, to situations where they neglect to undertake reasonable measures to control the risks emanating from robots has been suggested. In this regard, see Sabine Gless, Emily Silverman and Thomas Weigend, 'If robots cause harm, who is to blame? Self-driving cars and criminal liability' (2016) 19 New Criminal Law Review 412.

using of random/stochastic approaches to expand their problem-solving capability and eventually maintain their learning ability (autonomy, again). The former triggers the issue of liability of the data producer or of the network, feeding the data to the AI, which can have problems affecting the AI decision making process severely. The latter, might trigger eventual liability of the programmer or of the producer embedding the AI.[6]

Thus, unpredictable behaviours are the real game changer in the issue of liability attached to the use of AI. A new notion (Emergence Intelligence) has been created to refer to the ability of the system to engage in unpredictable behaviours still deemed useful despite (or because of) their unpredictability.[7]

Moreover, AI to remain 'intelligent' should maintain some form of continuous learning that constantly updates and changes the algorithm. For liability allocation this requires to target the 'right' algorithm involved in the specific decision-making process under scrutiny for liability. It requires accountability in the computational technical sense and a holistic approach.[8]

An example (the so-called 'runway trolley' scenario) in the AI setting (e.g. of biased decisions) will illustrate it. Suppose an autonomous vehicle is presented with the following alternatives: 1) to divert its path to save a busload of schoolchildren, but kill the vehicle's occupants in the process by colliding with a tree; or 2) to save the occupant, but let all the children

---

6 The consequences are manifold. For instance, KC Webb suggests that manufacturers must seek liability protection via legislation, leading the way to establish a national insurance fund, and develop training modules for buyers as part of purchase and lease agreements. KC Webb, 'Products liability and autonomous vehicles: who's driving whom' (2017) 23 Richmond Journal of Law and Technology 1.

7 Ryan Calo, 'Robotics and the Lessons of Cyberlaw' (2015) 103 California Law Review 513.

8 Jack Balkin clearly illustrates this need of a holistic approach: 'The laws of robotics that we need in our age are laws that control and direct human beings who create, design, and employ robots, AI agents, and algorithms. And because algorithms without data are empty, these are also the laws that control the collection, collation, use, distribution and sale of the data that make these algorithms work. […] So the laws we need are obligations of fair dealing, non-manipulation, and non-domination between those who make and use the algorithms and those who are governed by them'. Jack M. Balkin, 'The Three Laws of Robotics in the Age of Big Data' (2017) Yale Law School Faculty Scholarship Series 5159. Available at <http://digitalcommons.law.yale.edu/fss_papers/5159> accessed 28 July 2018.

die. Assume further that the AI 'driving' the vehicle learned what to do and decides autonomously its course.

Suppose something its machine-learned /was taught (the alternatives have different implications for liability we cannot discuss here in detail) is to reduce social cost (also in terms of compensation needed for victims following actual tort rules). Accordingly, it decides to kill the children because the loss of earnings of the vehicle occupant outweighs compensation for the children.

Under this exemplary scenario, we should ask ourselves whether there is a cause of action. And against whom? The non-existent programmer? The user of the car 'selecting' the tort interpretation to follow? The producer? Would it be different if a programmer actually taught (programmed) the rules to follow or the AI just learned them by 'reading' the case law provided for its training? Last but not least, we should investigate whether or not the algorithm should be tweaked to solve the problem? With which implications and by whom?

III. First policy implications

These very preliminary and simplified considerations already highlight, for law and policymakers, some key changes that stem from algorithmic decision-making used in AI and are extremely relevant for defining liability regimes:

1. Choices of algorithms embed specific policy choices and algorithms automatically enforce them.
2. Yet algorithms can permit direct accountability to the public, or other third parties (e.g. users, bystanders, producers, etc.) using the technological and legal frameworks already in place by providing analysable evidence of decision making processes, at least under some circumstances that can be programmed for scrutiny in an adjudication process;
3. Full transparency does not automatically help deciding liability issues;
4. Full Transparency is neither sufficient nor necessary (solving many IP and competition roadblocks in regulating AI related liability) for accountability and does not automatically help deciding liability issues;
5. Algorithms learn/are taught to (and in any event, they do) enforce their own interpretations of law and standards. It is for example the case,

stressed in literature[9] for algorithms automatically and autonomously 'enforcing' IP rights by sending notifications to (assumed) infringers.

6. Thus:
   a) In the absence of a clear regulatory framework algorithms might effectively advance the intermediaries' own interpretation of legal norms (instead of the legislator intended ones);
   b) The individual costs of reversing these interpretations might be so high (as opposed to their actual merit) to entirely beg the policy goals of the regulation.
   c) They risk self-setting the standard to abide to (requiring cautions in using soft law and self-regulation mechanisms).
   d) They might learn and reinforce wrong standards (e.g. with the task of minimising the costs of an accident, in the trolley car scenario they might decide to sacrifice a higher income individual instead of a lower income one or *vice versa*).

Regulators should take into account all these issues in reviewing the current debate on AI liability, a debate that might be driven too much by the examples of driverless vehicles since many valuable artificial intelligent agents are already employed in many fields and fully autonomous vehicles are not. Notwithstanding, surveying the autonomous cars liability debate enables to set the scenario for a more thoughtful intervention that considers liability in a wider legal and policy framework, a holistic approach that considers competition, IP, data protection, etc. After all in the common perception autonomous vehicles represent a sort of archetype of autonomous products in the meaning we clarified earlier.

IV. The debate on Autonomous vehicles as an example of an insufficient path.

Paradoxically as it might be, we do not have fully autonomous car traffic yet, but there is already extensive literature addressing the issue of liability in case of accidents caused by autonomous vehicles. We can quickly summarise the various positions with their arguments as follows.

---

9 Maayan Perel and Niva Elkin-Koren, 'Accountability in Algorithmic Copyright Enforcement' (2016) 19 Stanford Technology Law Review 473.

A number of authors sustain products liability or autonomous liability regimes based on strict liability. For instance, David Vladeck[10] claims that in the case liability concerns with the vehicle are the result of human (but not driver) errors, the products liability settled principles should be adopted to govern artificial intelligent machines, such as driverless cars. In other words, for the classical product liability defects (design, manufacturing and information defects failing to instruct humans on the safe and appropriate use of the product) a plain application of products liability rules would suffice, since there would not be a justification for treating even autonomous thinking machines differently than any other machine or tool a human may use; except, perhaps, holding them to a higher standard of care.[11]

In the case these machines cause injury in ways wholly untraceable and unattributable to the hand of man, i.e., that cannot fairly be attributed to a design, manufacturing, or programming defect, and where even an inference of defect may be hard to justify, the only feasible approach (it is suggested) would be to infer a defect of some kind on the theory that the accident itself is proof of defect, even if there is compelling evidence that cuts against a defect theory, as a simply restatement of *res ipsa loquitur* theory.

Leaving aside the problems and objections an interpretative approach fully based on *res ipsa loquitur* arguments would rise, the shortcomings of this reading appear evident by merely considering the role machine learning has in developing the driving algorithms and the implication the inability to understand the reasoning of them when they are black-box ones (unreadable to humans)[12].

Other authors[13] claim that autonomous vehicles should be treated like non-automobile products that have similar features, like elevators or autopilot technology (such as autopilot in ships and aeroplanes). Therefore, automobile liability regime should not be applicable to autonomous vehicles because they are 'too far removed from current automobiles both in

10 David C Vladeck, 'Machines without principals: liability rules and artificial intelligence' (2014) 89 Washington Law Review 117.

11 ibid 127.

12 As mentioned, when only their inputs and outputs are known it is not possible to have any knowledge of its internal workings even for its producer.

13 Jeffrey R Zohn, 'When robots attack: how should the Law handle self-driving cars that cause damages' (2015) 2 Journal of Law, Technology and Policy 461.

function and likely cause of injury'.[14] In the first case, 'autonomous cars can follow the evolution of elevator liability by beginning with a more standard negligent principle and then evolving to a more stringent and higher standard on the manufacturers over time and as the product continues to improve'.[15] In the case of autopilot technology, liability is attached to the manufacturer unless there has been negligence by the user.[16] Here the main shortcoming relates to the notion of 'autonomy' we discussed earlier. An elevator or a self-pilot in an aeroplane does not enjoy the same degrees of liberty and unpredictability a truly autonomous AI does.

A different approach[17] considers AI as software arguing that liability rules should follow those applicable to software because 'we cannot treat Al as legal entity […] since they do not bear own consciousness nor have independent property.'[18] However, 'Al can make a valid contract or do other legally binding declarations through its individual decisions but these are binding the represented person.'[19] By considering AI as software, these authors analyse the different paths in terms of liability rules and attribution of responsibility arising from this premise. Reference, for example, is made to Giovanni Sartor's arguments to explain that in the case of software we can count on more legal entities from the viewpoint of legal responsibility. According to Sartor, parts of the software agent could have separated legal fate. If the agent contains copyright protected software (as in most situations), the author could bear liability for programming mistakes. If the agent contains some kind of database, then the producer of the database could bear the liability for database mistakes. If the agent processes personal data, – from a data protection point of view – the data controller is responsible for legitimate data processing. If the agent is being operated by a certain user for own purposes, the user is liable for its operations. According to Sartor, since the usage of the software agent could have such aspects on which the given legal entity cannot exercise control, therefore the operator should not be held liable for such damages.

---

14 ibid 484.
15 ibid 483.
16 ibid 481.
17 D Eszteri, 'Liability and damages caused by artificial intelligence – with a short outlook to online games' (2015) 153 Studia Iuridica Auctoritate Universitatis Pecs Publicata 57.
18 ibid 65.
19 ibid 66.

For example: the user should not be liable for damages caused by Al software when the wrongful act originates from the programming mistakes of the software and the user is not allowed to access the source code or either to decrypt it.[20]

These authors clearly argue for a reasoned extension of existing legal rules and Sartor illustrates how different legal rules can deal with various instances of AI related liability paving the way to what we could call a multilayered approach.

A number of authors advocate a more radical shift from fault liability to strict liability based on defective products theories. For instance, some authors,[21] analysing the liability for damages caused by autonomous vehicles according to Belgian Law, argue that the fault-based regime entails several problems because: 1) it is unlikely that a victim will be able to prove that the user of the vehicle acted negligently because of the unpredictability of software systems and the rigid interaction with users make it harder to assess the reasonable foreseeability and avoidability of the harm, which is an essential element for a negligence claim[22]; and 2) it is unlikely that a human would be able to avoid damages that a computer cannot.

Accordingly, the solution would be to evolve from fault-based to strict liability in traffic-related matters. In this regard, autonomous vehicles that caused damage due to a dysfunction in their software or hardware will be seen as a defective product and claims could be filed either against the manufacturer of the vehicle or the software producer depending on the factual situation. In any case, according to Belgian Law it is unsure whether the software producer could be held liable. Some improvements on defective products regime would remain necessary.

Other authors propose a different strict liability test (reasonable car standard test) adapted to autonomous vehicles. This test should be applicable in the strict liability products regime and would hold a car manufacturer liable only when the car does not act in a way that another reasonable autonomous vehicle would act. Accordingly, the test would be 'determining how a reasonable AV would act and comparing it to an allegedly deviant AV would be far less invasive and expensive for the parties than liti-

---

20 Giovanni Sartor, 'Cognitive automata and the law: electronic contracting and the intentionality of software agents' (2009) 17 Artificial Intelligence and Law 253.

21 J De Bruyne and J Tangue, 'Liability for Damage Caused by Autonomous Vehicles: a Belgian Perspective' (2017) 8 Journal of European Tort Law 324.

22 ibid 346 and 371.

gating whether a safer alternative design would be implemented by comparing lines of computer code.'[23]

Feasible or not such an approach illustrates again the need to move from existing liability regimes (product liability remaining one among the others) and adapt them to the novelties brought about by AI. Yet, the key issues on the adaptation of the existing liability regimes (or the creation of new ones for what it matters) are not fully tackled: above all, the way the different regimes should interact among them; what we could call the problem of splitting the bill among the different potential tortfeasors.

Finally, yet importantly, there are authors who propose strict liability rules in the field of AI, by analogy with a party's responsibility for the behaviour of animals, children, employees or even ultra-hazardous activity. These authors' approaches vary extensively. For instance, some consider AI systems as a tool using the general principle contained in article 12 of the United Nations Convention on the Use of Electronic Communications in International Contracts. According to this principle, 'the principal of a tool is responsible for the results obtained by the use of that tool since the tool has no independent volition of its own'[24]. Accordingly, they propose strict liability rules by analogy with a party's responsibility for the behaviour of children and employees (vicarious liability). Applying this regime for the case of AI 'means that liability for the actions of AI should rest with their owners or the users.'[25] Therefore, liability is imposed on the person, not because of his/her own wrongful act, but due to his/her relationship with the tortfeasor AI.[26]

In the case the AI is a source of danger they suggest treating AI liability as that applicable to certain types of activities associated with greater danger to others. Therefore, in these situations, the AI developer should be held liable for creating this greater source of danger, and, liability arises without fault by using the '*Cuius commoda eius et incommoda*' theory, 'which means that a person engaged in dangerous activities that are profitable and useful to the society should compensate for damage caused to society from the profit gained'.[27] This person can be the AI producer or

---

23 KC Webb, 'Products liability and autonomous vehicles: who's driving whom' (2017) 23 Richmond Journal of Law and Technology 1.
24 Bruyne and Tangue (n 20) 386.
25 Bruyne and Tangue (n 20) 385.
26 Bruyne and Tangue (n 20) 387.
27 Bruyne and Tangue (n 20) 386.

programmer, who should be required to insure against civil liability as a guarantee for their hazardous activities. Another possibility would be to use the Common Enterprise Doctrine, adapted to a new strict liability regime.[28]

These final examples are a clear illustration of the potentials and limits of mere extending the existing legal rules to the unsettling elements of AI related liability. A deeper analysis of other factual examples, for instance referring to robotics or IoT[29] demonstrates that any liability regime (extended from existing rules or created anew) is but one (insufficient) layer in the multilayered liability environment the complexity of these new technologies require.

We will briefly argue that a multilayered liability system based on accountability would prove more apt to solve the emerging legal issues and will help solving the allocation of costs among the possible liable entities (the split the bill – among tortfeasors-problem).

V. Towards a multilayered liability approach based on accountability

Before providing more details on the role of the concept of accountability, it is useful to summarise the most fundamental questions not answered by interventions on applying liability rules to Robotics, IoT and AI in general.

How to deal with the issues raised by AI (e.g. black boxes issues and the 'defects' produced by machine learning)? How to allocate costs among the various players in the development, deployment, use of AI? (e.g. algorithm developers, products producers, connection providers, data providers, data brokers, users, the State, bystanders, infrastructure maintenance, sensor producers, …).

Even when duly adapted, traditional tort liability rules are ineffective in sorting the allocation of liability and costs related to AI (recourse, contribution, joint and several liability, etc.). For distributing the costs among

28 Bruyne and Tangue (n 20) 387.

29 See Goldman Sachs, 'The Internet of Things: Making sense of the next megatrend' (2014) *Iot Primer*, available at <https://www.goldmansachs.com/our-thinking/outlook/internet-of-things/iot-report.pdf> accessed 28 July 2018; Mauricio Paez and Mike La Marca, 'The Internet of Things: Emerging Legal Issues for Businesses' (2016) 43 Northern Kentucky Law Review 29.

the mentioned players, the public and the insurance market (to address the 'split the bill problem') accountability would still be needed to allocate effectively liability. Indeed, whatever liability basis is chosen (fault, strict liability, vicarious liability, no-fault, funds, AI direct liability, etc.) legal and computational accountability mechanisms effectively help splitting the bill in 'automated' and 'verifiable' ways at virtual no costs.

A simple example will illustrate this point. Any selected liability regime would set a 'required standard of conduct' that would request to establish a number of causal connections between the actions/omissions targeted as it triggers of liability to the 'required standard of conduct'. However, most of the time, it would be necessary to ascertain these causal links notions of computational accountability applied to the legal issues in discussion.

What we are suggesting is that AI requires a gradual layered approach to liability grounded on accountability principles (already embedded in the EU legal system). AI requires the use of technology itself to unfold a multilayered accountable liability system and solve the 'splitting the bill problem'.

The need to embed technologies in liability rules and not only attach liability to technology is illustrated by the example of the required standard of conduct. While the deviation from the standard conduct by humans is assessable by humans, the deviation from the standard of conduct by AI is assessable only with the help of 'technologies' with the characteristics required by the accountability principle.

The almost automatic verification of the causal reasons for AI's actions might then trigger different layers of the multilayered liability system.For example, verification of full compliance with the set of rules coded in the AI and made available could trigger the application of compensation funds or a reversal of the burden of proof in a given liability regime, channelling liability on different individuals, a stricter or more lenient regulatory system, various kinds of insurance, the ability to contract around liability, the shift of liability on users, and so on.

In a word, any path suggested by the European Parliament resolution on liability of robots[30] or the European Commission on building a European data economy requires this approach.

---

30 European Parliament Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

VI. For a theory of layered liability: accountable multilevel liability

As the brief discussion of the literature on driverless car liability illustrates, as a baseline in defining the liability regimes applicable to Robotics, IoT and their AI we should acknowledge that we do have liability rules and principles theoretically applicable to them. As illustrated, the debate swings between a choice of fault, no-fault, strict liability, vicarious liability, on the one hand, and the attribution of liability on the AI agent itself on the other. Normally, the application of existing rules and principles is only a matter of interpretation and asks simple yet relevant questions jurists are used to answer using the traditional hermeneutical tools. Reinterpretation, however, presents its own limits and certainly cannot cover every facet of the changes fostered by the tackled technologies.

In relation to liability applicable to AI and autonomous systems, there is a frequent criticism pointing to the fact that the use of AI and autonomous systems could create responsibility gaps rendering impossible to attribute legal liability to anyone for harms caused by the autonomous operation of these technologies. This is the case for instance because neither designers nor users would be at fault with regard to errors and harms which are unknown and undiscoverable at the time the product is placed on the market and therefore could not be anticipated or remedied given the existing technological knowledge.[31] Nevertheless, legal scholars have been trying to overcome this gap by applying and adapting traditional civil liability rules to damages caused by artificial intelligence. Accordingly, authors have proposed different solutions for the prospective scenarios they envisage. These solutions are usually based either on the strict liability regime of party's responsibility for the behaviour of people or animals under their responsibility or on the strict liability regime for products, abnormally dangerous activities or wild animals, both of them with some adaptations. Moreover, traditional civil liability rules also provide for special regimes that attribute responsibility to persons not actually responsible for perpetrating a wrongful act due to their relationship with the perpetrator, as in the case of liability of owners of guardians or vicarious liability.

As anticipated, among the proposal on the floor seriously considered at the EU level is to attribute liability directly to the AI (*rectius*: in the pro-

31 Giovanni Sartor and Andrea Omicini, 'The autonomy of technological systems and responsibilities for their use' in Nehal Bhuta et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016).

posal to the robot using the AI).[32] However, also this proposal does not answer the key issues and cannot work properly. For reasons we cannot explain here liability of the autonomous agent systems might create more problems than solving ones.

The idea that AI can have autonomy and should trigger the liability of the AI itself might sound naïve or futuristic. However, for centuries we have been attributing liability on non-humans, shielding humans consequently from liability. Liability of legal persons (e.g. corporations) is a suitable example of this approach. Yet, the similitude would work upon the condition that we confer a sufficient patrimony to the AI to fulfil eventual compensation duties. However, it is easy to see the actual risk of creating AI entities with limited assets to the exclusive aim of reducing liability for their creators/users.

As anticipated, the borderlines between automated and autonomous[33] systems are progressively blurring for several reasons. It is worth remembering that even automated systems increasingly employ AI with embedding machine learning that makes opaque the way the AI operates and difficult the understanding of the role data (both training and sensor data) play in the operation itself. Moreover, they can operate in unstructured environments, making them dependent on sensor data and information, and using random/stochastic approaches to expand their problem-solving capability and eventually maintaining their learning ability. Here, again, autonomy is not of a moral character but rather a degree of relative absence of control triggering various degrees of liability in using the 'autonomous' AI.

As already acknowledged in most jurisdictions the principles of *respondeat superior*, *ubi commoda et eius incommoda* in vicarious liability see rules limiting (e.g. defences or recourse actions) liability of the *prima facie* responsible agent (employers, users, producers). This happens, for instance, when the collaborator violates the instructions or rules set by the person who avails herself of the collaborator. We might be tempted to consider AI (at least one showing autonomous decision-making) as we consider collaborators. Yet, the opacity of the AI decision-making process or its biased evolution requires the use of technology itself to guarantee a

---

32 European Parliament Resolution (n 29.)

33 For a caution not to confuse autonomy in philosophical and in robotic terms see Noel Sharkey, 'Saying "No!" to Lethal Autonomous Targeting' (2010) 9 Journal of Military Ethics 369.

correct assessment of the deviation from the expected/commanded conduct.

In other words, while the deviation from the standard of conduct by humans is assessable by humans, the deviation from the standard of conduct by AI is assessable only with the help of software with the characteristics required by the accountability principle. It is technically possible to a significant extent that computer science itself enables the almost automatic verification of the causal reasons for AI's actions in a number of instances. This might then trigger different layers of a multilevel liability system. For example, full compliance with the set of rules coded and made available could trigger the application of compensation funds, a reversal of the burden of proof, channelling liability on different individuals…

In addition, the same mechanisms, guaranteed by making AI accountable, would help apportioning the costs among various stakeholders and across different liability regimes that would probably coexist at least for some time.

VII. Blending liability and accountability for selecting AI liability regimes

It is now time to understand better the meaning and role of accountability. It has precise (although varied) meanings in both computer science and law.

In law 'Accountability refers to the extent to which decision-makers are expected to justify their choices to those affected by these choices, be held answerable for their actions, and be held responsible for their failures and wrongdoings'.[34] Thus, relevant literature already acknowledged that accountability indicates liability to account for and answer for one's conduct, the obligation to provide a satisfactory answer to an external oversight

34 Maayan Perel and Niva Elkin-Koren (n 8). See also Michael D Dowdle, 'Public Accountability: Conceptual, Historical, and Epistemic Mappings' in Michael D Dowdle (ed), *Public accountability: Designs, dilemmas and experiences* (Cambridge University Press 2006) ('persons with public responsibilities should be answerable to "the people"' for the performance of their duties.'); Danielle Keats Citron and Frank Pasquale, 'Network Accountability for the Domestic Intelligence Apparatus' (2011) 62 Hastings L.J.1441 (focusing on accountability as a measure to cure the problems generated by the growing use of Fusion Centers); Tal Zarsky, 'Transparent Predictions' (2013) U. Ill. L. Rev. 1530.

agent. A notion well known also in our legal systems as clearly exemplified by the EU General Data Protection Regulation.

This notion of accountability has a counterpart in computational disciplines. Here it is defined as 'set of mechanisms, practices and attributes that sum to a governance structure […] for processing, storing, sharing, deleting and otherwise using [personal and/or confidential] data according to contractual and legal requirements. Accountability involves committing to legal and ethical obligations, policies, procedures and mechanism, explaining and demonstrating ethical implementation to internal and external stakeholders and remedying any failure to act properly'.[35] Surprisingly enough computer scientists are more aware of the links between computational accountability and law than jurists are.

As already illustrated in literature[36] accountability can have a procedural and a substantive meaning. In the former, it indicates a sort of procedural regularity ensuring each individual that the same procedure is applied to them and that the procedure was not designed in a way that disadvantages them specifically. In substantive terms, accountability imposes that the policy furthers fundamental goals or principles addressing legal and ethical questions: does the rule implemented correspond to moral, legal, and ethical criteria. Is it actual operation faithful to these substantive choices?

Along these premises, we can develop a threefold notion of accountability for AI (a notion which is already implemented in the GDPR, for instance):

1. The obligation to report back to someone (*rendre le compte*) to show how responsibility is exercised and making this verifiable;
2. Enabling the 'Audience' (stakeholders, victims, authorities) to interrogate and question the accountable entity thus producing 'their own accounts';
3. Providing for various levels of 'Sanctions' when a violation of accountability occurs (inaction or bad actions).

This threefold notion of accountability requires 'demonstration' mechanisms and criteria to assess them: enabling more than correlation in compu-

---

35 IEEE, *A Glossary for Discussion of Ethics of Autonomous and Intelligent Systems*, available at <https://standards.ieee.org/develop/indconn/ec/eadv 2_glossary.pdf> accessed 8 August 2018.

36 Joshua A Kroll et al, 'Accountable Algorithms' (2017) 165 U Pa L Rev 633.

tation, giving certainty to legal standards. Accountability is an *ex-post* instrument that requires *ex ante* actions (an already emerging regulatory framework) to enable the provision of evidence in an adjudication process.

Finally, accountability enables different layers of liability: reversal of the burden of proof, compulsory insurance, funds, regulatory constraints, criminal sanctions. These are made more easily applicable and scalable using algorithm (procedural and substantive) accountability for liability purposes.

It is important to note that technology is today able to offer a number of tools to implement effectively algorithm (procedural and substantive) accountability for liability purposes.[37] Although we do not have here the possibility to illustrate it in detail, the accountability principle coupled with a proper use of technology enable to assess and show, for instance, that the deviation from the expected conduct is triggered by the specific use to which the AI has been put in use for or that the deviation from the expected conduct is triggered by the dataset originally used in the machine training of the AI. Intuitively, this ability to allocate factual causality would prove essential in the allocation of liability among potential multiple tortfeasors/stakeholders. It is a possible way to solve the split the bill problem.

Of course, we are aware that there are technical restrictions to the use of such technologies. For instance, the potential interconnectedness of algorithms and of algorithmic decisions seriously restricts the means of algorithmic decision-makers to give an account of the decisions they make.

---

37 Among these techniques are Software verification (on this subject see Jean Souyris et al, 'Formal verification of avionics software products' in Ana Cavalcanti and Dennis Dams (eds), *FM 2009: Formal Methods* (Springer 2009) 532, Norbert Völker and Bernd J Krämer, 'Automated verification of function block-based industrial control systems' (2002) 42 *Science of Computer Programming* 101; Daniel Halperin et al, 'Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defences' (2008) *IEEE Symposium on Security and Privacy*, 129; Karl Koscher et al, 'Experimental security analysis of a modern automobile' (2010) *IEEE Symposium on Security and Privacy* 447; Stephen Checkoway et al, 'Comprehensive experimental analyses of automotive attack surfaces' (2011) *20th USENIX Security Symposium* 77) algorithms explanations to make them interpretable (See Cynthia Rudin, 'Algorithms for interpretable machine learning' (2014) *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge discovery and data mining* 333) Cryptographic commitments, zero-knowledge proofs, fair random choices (on these techniques see again Kroll et al, ibid).

However, we can still assert with confidence that the basis of the debate surrounding liability regimes for the tackled new technologies have changed and are partially determined by the technologies they aim at ruling. Time is ripe to discuss how to unfold a multilayered accountability-based liability system using both:

a) existing and forthcoming experience/technology for accountability and
b) the existing/emerging regulatory framework at the EU level (e.g. GDPR is a clear example[38]).

Here we can only state that AI requires a gradual layered approach to liability grounded on accountability principles (already embedded in the EU legal system). In addition, we cannot illustrate here fully that AI requires the use of technology itself to unfold a multilayered accountable liability system and solve the 'splitting the bill problem'.

To conclude, let's have a closer look into this last issue.

Even when duly adapted, traditional tort liability rules are ineffective in sorting the allocation of liability and costs related to AI (recourse, contribution, joint and several liability, etc.); distributing the costs among the mentioned players, the public and with the insurance market. To the contrary, whatever liability basis is chosen (fault, strict liability, vicarious liability, no-fault, funds, etc.) predictable legal and computational accountability mechanisms effectively help splitting the bill in 'automated' and 'verifiable' ways at virtual no costs. A number of the causal connections and the correspondence to the 'required standard of conduct' can rely on legal computational accountability.

Note also that if the issue of apportioning the financial costs of liability is left to insurance companies by way of redress actions, the need to have objective criteria to redistribute the costs remains and can effectively rely on the accountability notion we are briefly introducing.

The link between the accountability principles and any liability regime remains essential.

---

38 'Accountability is an important area and an explicit requirement under the GDPR.' (WP29 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679', adopted on 3 October 2017, as last Revised and Adopted on 6 February 2018, at 29).