Acta Universitatis Sapientiae

Legal Studies

Volume 8, Number 2, 2019

Sapientia Hungarian University of Transylvania
Scientia Publishing House

The Law and Artificial Intelligence project and conference organised by the Sapientia University, Cluj-Napoca (May 9-10, 2019) have been realised with support from the Hungarian Ministry of Justice.

Contents

Acta Univ. Sapientiae, Legal Studies, **8**, 2 (2019) 151–165

DOI: 10.47745/AUSLEG.2019.8.2.01

The Double Side of Artificial Intelligence in the Public Sector

Isabella Alberti
PhD student in administrative law
University of Turin, Turin
E-mail: isabella.alberti@unito.it

Abstract. The introduction of artificial intelligence in the public sector seems to be both a positive and negative development. On the one hand, artificial intelligence could improve the efficiency of public bodies due to the acceleration of the decision-making process, especially for repetitive procedures to free up public servants. Big data analysis, artificial intelligence, and the Internet of Things applied to the public sector could allow the reshaping of public service delivery. This is so, on the one hand, because data becomes a 'piece of reality' and, therefore, the aggregate analysis of data gives a realistic and objective picture of the current society. On the other hand, some concerns arise when artificial intelligence dissociates civil servants from the recipients of their services or affects the rights of these recipients. Scholars are called upon to reflect on the nature of artificial intelligence to overcome obstacles related to the 'black box' nature of its functioning and to better implement it in the public sector field. Legal rules and principles in the administrative decision-making process play a crucial role as they risk hindering the development of artificial intelligence in the public sector, as the Italian case-law highlights.

Keywords: artificial intelligence, public sector, evidence-based regulation, decision-making process

1. Introduction

This contribution aims to analyse the impact of artificial intelligence on the public sector due to the many benefits to the community and also its impact on individual rights and freedoms.

The introduction of artificial intelligence in the public sector seems to be both a positive and negative development. It is clear that artificial intelligence could improve the efficiency of public bodies due to the acceleration of the process of decision making, especially for standardized and repetitive decisions, freeing up

human resources. Furthermore, big data analysis, artificial intelligence, and the Internet of Things applied to the public sector could allow the reshaping of the delivery of public services. The blockchain, for example, could strengthen the transparency and traceability of decisions that involve several public agencies.

Generally, new technologies and the high amount of data represent an efficient tool to target problems more effectively and in a timelier manner by harnessing the data collected on different social groups. Thanks to the knowledge gained in the field of network science, reality could be explained as a space constituted of *interconnections* (links or vertices) and *nodes* (hubs), resulting in a repetitive and almost universal behavioural scheme.

The high degree of digitalization in the private and public sector, the computational power of computers, and the ubiquity of wirelessly interconnected devices – which can capture, store, and transfer data to computer servers – definitely change how reality can be represented: data becomes a 'piece of reality' and, therefore, the aggregate analysis of data provides a realistic and objective image of current society.

Consequently, digital tools provide the public administration with the knowledge necessary to act in a new way, strictly focused on achieving various public interests. It results in interconnectedness at different levels of government.

Nevertheless, some concerns arise when artificial intelligence substitutes civil servants and affects public service recipients' rights. As in the private sector, where algorithms may decide if someone can receive a loan, the same could happen in the public sector (for example, to decide if someone could receive social benefits or subsidies). However, the private and public sectors are different in scope: the first one tries to maximize individual profits, while the second one must pursue public interest – without unreasonable discrimination. Furthermore, every national and European framework possesses some specific principles which public bodies should adhere to, such as *transparency* and *impartiality*, the obligation to provide reasoned decisions, the right to be heard or to participate, to resolve the requests of recipients (a right to due process and good administration). Related to these public priorities and principles, Article 22 of the General Data Protection Regulation (GDPR) – which came into force in all Member States of the European Union at the end of May 2018 – bans all decisions adopted in an automatic way that affect data subjects (with only three specific exceptions).

Consequently, in the administrative decision-making process, legal rules and principles seem to limit the use of artificial intelligence, as the Italian case-law highlights. Scholars must reflect on the nature of artificial intelligence to overcome obstacles related to the so-called 'black box' nature and to fit them better in the public sector field. For example, scholars should decide if artificial intelligence entities should be considered merely as tools or, indeed, as artificial civil servants. This preliminary choice is relevant to building a framework that could be observant of both the core principles of administrative decision making and the rule of law.

The positive effects of artificial intelligence are clear and extremely beneficial for improving innovation and competitiveness in the public sector. However, we cannot ignore its negative effects. For this reason, scholars and policymakers must reflect on a safe and respectful framework in which artificial intelligence could be legally used also in the public sector. In this process, lawyers especially should adopt an interdisciplinary approach that could allow them to better understand the fourth industrial revolution which we are experiencing, the formation of a digital society.

2. Data and Artificial Intelligence: How Does the Delivery of Public Services Change?

In the public sector, the Italian process of innovation began many years ago, albeit slowly. In the 1990s, the first attempt of digitization of public documents and the promotion of digital tools were made. After several years of effort, concerns grew because of the difficulties associated with giving legal value to digital documents, and, consequently, the digitalization of public administration proceeded slowly.

The advent of digital revolution, coupled with the growth of computational power – two key features of the Fourth Industrial Revolution –, the interconnection of digital devices triggered the digitalization and automatization of the public sector. Not only did new tools for providing public services and for facilitating the relationship between public bodies and citizens become available but also a new idea of public bodies as makers of knowledge emerged, based on quickly collected administrative data.[1] This had the consequence of making data available to third parties as they are considered drivers of innovation and economic development.

The advantages of these transformations are evident. Nowadays, an increasing number of public services are accessible online through mobile applications and websites; the communication between public bodies and citizens is easier and more digital-based. Information about traffic, weather, and so on is displayed, for example, on real-time dashboards. Every day, the public sector could collect a large amount of personal and non-personal data, known as administrative data, as well as urban data on city infrastructures and utilities (i.e. traffic, public transports, etc.). All data are archived in different databases. Unfortunately, in the Italian

1 Indeed, throughout the centuries, public bodies just collected data about cities and their citizens, but they were based on relatively limited samples, time- and space-specific, with the restricted number of variables. They have been defined as *small data*, which refer to data captured with questionnaire surveys, case studies, city audits, interviews, focus groups, national censuses, and government records. 'Small data' is characteristically limited and outdated because of inadequate tools for capturing and analysing it.

context, there's a so-called 'data silo approach' in effect, according to which each public body builds a specific database to satisfy its own needs. This approach is patently counterproductive in the era of artificial intelligence because of the way new technological tools function. Artificial intelligence needs to be trained with a big amount of data to offer some decision-making proposals and to highlight useful patterns.

In the light of this, the Italian Normative Code (d.lgs. 82/2005) and the efforts of the Digital Team and the Agency for Digital Italy are trying to transfer all administrative data to a public platform – called the Data & Analytics Framework (DAF) – where all local and national public bodies may access and use data for their decision-making processes.[2] This transition is meant to achieve the 'data lake' paradigm by implementing the idea that sharing data is now the most important challenge that all public bodies must deal with. To collect and to analyse a big amount of data is the first step to augmenting the ability of public bodies to gain knowledge and to inform the policy. The DAF platform tries to enhance and simplify the interoperability and the exchange of public data through public bodies to obtain more information as well as to standardize and to open data to third parties.[3] This approach makes the delivery of public services more efficient because it improves precision and rationality while reducing waste, based on a data-driven approach.[4] Until now, public bodies have always acted with an *ex-ante* programmed approach, but now the interconnection among big data, artificial intelligence, and the Internet of Things is predicted to allow public bodies to deliver public services efficiently for policy and governance purposes. Data-driven regulation based on the interconnection of data from different sources and the application of artificial intelligence, such as predictive algorithms, could propose to public bodies some solutions to achieve the most important goal of getting more for less. It represents an evident application of the potential of innovation that could move up the performance curve of public sector action.

Evidence-based methods build on trends and an *ex-ante* perspective. The information available at the time of regulation is paramount. As a result, evidence-based law-making processes are growing thanks to indicators, which capture real-time data and translate it on dashboard graphs, which provide detailed information about these indicators in a human-accessible form. For example, city performance

2 For more information, see: https://teamdigitale.governo.it/it/projects/daf.htm. See also: Tresca 2018.

3 A new directive on the open data paradigm and the regime of public data is set to come into force: Directive (UE) 2019/1024 of the European Parliament and Council of 20 June 2019. This act will substitute the Public Sector Information Directive (UE) 2013/37 of the European Parliament and Council of 26 June 2013.

4 On the benefits of data-driven regulation, see: Di Porto–Rangone 2013; Borgogno–Colangelo 2019.

may be measured in this way. This revolution allows public bodies to capture 'the desire to reform the public sector management of city services to make them more efficient, effective, transparent and value for money, combined with citizen and funder demands'.[5]

Data-driven regulation and artificial intelligence could enhance the efficiency and performance in fields specific to the public sector: the goal is to customize and personalize the action as well as to make more informed decisions and act on them[6] to cut some costs, to 'do more with less'.[7] The impact of artificial intelligence can be visible also from the civil servants' side: many repetitive tasks or easy ones could be completed by the technologies to free up servants and to re-address their activities to tasks requiring more use of human discretion: activities such as opening e-mails and attachments, filling in a form, scraping data from the web, or extracting structured data from documents could be done by bots in place of civil servants.[8] Moreover, artificial intelligence could be used in other different ways, particularly for simplifying the relationship between public bodies and citizens: for example, the Municipality of Turin applies a voice assistant for helping users to understand how public offices fit better with their needs: in particular, some microphones and loudspeakers will be installed in the more crowded offices to answer to questions posed by users and help them to address these to the correct office.[9] Similarly, a vocal assistant, named EMMA, is used in the USA for helping foreigners to receive the right answers about immigration issues and services as well as to find information on the website.[10] Another example is represented in the languages and data process support, such as in the field of education, in which the platform named PIERINO (PIattaforma per l'Estrazione e il Recupero di INformazioni Online – Platform for Extraction and Recovery of Online Information) has helped the Ministry of Education in

5 Kitchin–Lauriault–Mcardle 2015. 8.

6 Some examples arise from the education and health sector. In the education field, some public schools are using artificial intelligence in order to provide a more specific and customized learning experience, individualized for each student: the *ratio* is that the traditional school methods are ineffective. Similar applications are found in the health sector, where thanks to artificial intelligence doctors could analyse and individualize more quickly and more accurately some responses to specific diseases related to each patient.

7 For this approach, see: Maciejewsky 2017. 120–135.

8 For many other examples, see: *AI-Augmented Government: Using Cognitive Technologies to Redesign Public Sector Work* 2017. It is also important to underline that there are four approaches to the issue of substituting the human factor: the *relieve* approach aims to free up public workers from repetitive tasks; the *split-up* approach suggests a collaboration between workers and machines, in particular giving repetitive tasks to a machine and supervisory control to a human; the *replace* approach that (quite) totally substitutes humans in doing simple tasks or in giving easy responses; finally, the *augment and extend* approach is based on complementary activities between machines and workers.

9 The Municipality of Turin has just trained some of the speakers and loudspeakers with the most frequently asked questions and answers.

10 For more information, see: https://www.uscis.gov/emma.

the analysis of thousands of answers provided by citizens on the state of (and desire about) school.[11]

3. Artificial Intelligence in Decision-Making Processes: Some Critical Examples

The transformation of government in an artificial administration is shifting from the professional treatment model to a bureaucracy rationality[12] model. In the first approach, a human professional could govern and manage singular situations related to unique recipients through a fair procedure that considers legality as well as the individual's situation. The bureaucracy rationality model is based on a repetitive and depersonalized approach in which the data entered and data-matching are more relevant than recipients' situations. On the one hand, the advantages coming from a bureaucratized and automatized procedure are evident.[13] Some challenges are emerging, in stark contrast to some administrative principles and the rule of law.[14] For example, the Australian government used automatization to verify recipients' income to avoid overpayment of social benefits. This tool, known as the Online Compliance Intervention (OCI) program,[15] was proposed to make the welfare system more sustainable and, especially, 'to recover money from people that deliberately seek to defraud the social welfare system as well as those who have simply inadvertently been overpaid'.[16] Before this program, the government was used to checking some variations of recipients' income through a strict collaboration between officials of the Department of Human Services and the Australian Taxation Office. After the introduction of the automated approach, officials lost direct control over the input of recipients' income variations. The responsibility of inputting data regarding income was transferred directly to the recipients. The verification process and the enforcement action were automated without human oversight. Every discrepancy between the Australian Tax Office

11 See more detailed information at: http://legacy.fbk.eu/it/news/fbk-collabora-con-il-miur-labuonascuola.

12 Here I refer to a model of administration of justice in which governmental bodies as well as tribunals are included. For more information, see: Mashaw 1983.

13 Some of these are explained here in Chapter 2.

14 On the challenges to rule of law in the digital era, see: Wright 2014. For a weekly update about this issue, see the website: https://binghamcentre.biicl.org/categories/digital-age; for a more profound analysis on the impact of the digital era on many aspects of the rule of law, see: *The Rule of Law on the Internet and in the Wider Digital World. Issue paper published by the Council of Europe Commissioner for Human Rights*, available at: https://rm.coe.int/16806da51c.

15 For general information on the case, see: Daly 2019.

16 Senate Standing Committee on Community Affairs (Australia), Design, scope, cost-benefit analysis, contracts awarded, and implementation associated with the Better Management of the Social Welfare System initiative, 21 June 2017, at para. 1.3.

records and the data provided by recipients was considered as a proof of undeclared or underreported income. Consequently and automatically, the letter was sent to invite recipients to update or correct the data: if a recipient was not able to update data, refused to do so, or just considered that the request was not well-founded, the enforcement action was begun without considering any kind of challenges of recipients,[17] and the request for purported debt was sent.

This event became known as the 'robodebt scandal'. Some politicians have paid attention to this, in particular because this kind of automation was characterized by 'disruption and impact to individual's lives',[18] while the public confidence in this system was reduced. Debt calculation was not transparent and was unavailable to the public. The 'robodebt scandal' highlighted that a high rate of digitalization without sufficient support provided to the users could create unfortunate consequences. In this case, for example, recipients – who cannot work with a digital platform or do not have the possibility to express their reasoning because of mental health issues or other kinds of social or economic disadvantages or just for the belief in the rightness of authority – lost the possibility to defend own rights. The main consequence is the discrimination of poor people that are in disadvantages and unable to contrast with the algorithm decision and, consequently, succumb to it.

After a politician denounced these disadvantages, proposals were addressed to the OCI program to overcome the lack of procedural fairness. For example, the provision of an independent review of internal and external debt collection practices as well as external scrutiny on the process were introduced.[19]

Another application of artificial intelligence in the administrative decision-making process is the HART (Harm Assessment Risk Tool), the AI-based technology that could help policy bodies[20] in deciding about custodial decisions in the jurisdiction of the Durham Constabulary. The program is based on over one thousand personal histories of people previously arrested and processed in the last five years and is used for predicting if a suspect could re-offend during the next two years. The automatic forecast about the risk and the recidivism of a person could be more useful for a quick decision-making process, but how it operates has been criticized. The program works out the new decision based on previous decisions and by cross-referencing some variables, such as the neighbourhood of origin, the age, the gender, the income, and so on, that are considered as predicting parameters. The risk of discrimination is therefore high because of the sensitive variables that are analysed. This program categorizes people to better individualize

17 E.g. difficulties in using the online portal or personal difficulties relating to family, job, or mental disease; more information is available in the above-cited report, in Chapter 4.

18 More information is available in the above-cited report, at para. 1.4.

19 More advice is available in the above-cited report, at para. 1.4.

20 For a broad analysis on artificial intelligence for criminal prevention, see: Bonfanti 2018, Babuta 2017, Scassa 2017.

the risk of a new crime. Even if it presents some advantages, practitioners should consider the necessity to update data and to insert correct data as well as seeing HART just as a tool, some sort of guidance and not as a decision maker.[21]

Similarly, in the American context, the Supreme Court of Wisconsin denied to a defendant named Eric Loomis a certain measure of freedom while under surveillance because of the Report released by the digital risk assessment tool called COMPAS[22] (Correctional Offender Management Profiling for Alternative Sanctions) which highlighted a high risk of danger and recidivism. As many commentators have underlined,[23] this decision seems to be discriminatory because it was based on a report generated on gender and race variables;[24] at the same time, the decision appears to be unfair because it is based on a proprietary software,[25] the functioning of which was consequently unknowable and unfit for external analysis.

Another critical application of artificial intelligence in the decision-making process is in university admission. One of the examples refers to the French context, where the portal Admission Post-Bac (APB) uses an algorithm for the selection and admission of students to the university and for creating a waiting list for available positions. The ranking in this waiting list is determined by use of the algorithm called Admission Post-Bac (APB), and it is based on the data available on students. In particular, each student is profiled[26] concerning the preferences expressed about universities, their school backgrounds, the postal code, and the family background.

If there are as many positions as students applying to them, there are no problems. Otherwise students are re-addressed to other, similar universities, without any possibilities to oppose the decision or to provide further data. Regarding this procedure, the *Commission nationale de l'informatique et des libertés* (CNIL – the National Commission of Information Technology and Liberties) determined some defects such as the absence of any information about the automatic collection of some personal data on family or grades at school.[27] Additionally, the Commission considered that the government has not provided sufficient information on the

21 For a proposal that considers the possibility to assess and to correct the algorithm's response, see: Oswald–Grace–Urwin–Barnes 2018. 244.

22 COMPAS is one algorithm used by the American police in order to individualize where crimes could be committed and prevent them as well as to provide personal information on suspects.

23 Supreme Court of Wisconsin, State of Wisconsin v Eric L. Loomis, Case No 2015AP157-CR, 5 April–13 July 2016; according to Carrer 2019; Simoncini–Suweis 2019. 93.

24 In 2016, a journalistic investigation demonstrated the racial biases in the software; see at: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

25 On the contrary, the Italian jurisprudence supports the accessibility to the software; see: TAR Lazio, sez. III-bis, sent. No 3769/2017; see also Pinotti 2019. 118–125.

26 The elements that could be analysed are individualized in the French *Code de l'Education*.

27 The French legal framework, according to the GDPR regulation, obliges public bodies to inform on automatized processes such as collection or managing personal data.

rejection of the request as well as the logic underpinning the decision, the weight assigned to any variables, or the rate of mistakes.

Consequently, the CNIL obliged the government to cease the use of this portal. Nevertheless, a new platform, called *Parcoursup*, has just replaced the old one. It fits the current legal framework[28] better, especially because it allows students to present some exceptional circumstances or an opportunity for discussion with the university, which has refused the student. Moreover, if there are no positions in the selected university, the platform addresses a proposal to the student to apply to another university, similar to the preselected one. The student has the opportunity to accept or to refuse[29] this proposal. Finally, the new platform respects the guarantees expressed in the GDPR framework.

Since 2016, the Italian Administrative Court handed down decisions on how artificial intelligence should be used in administrative procedures to respect the whole normative framework. The entry of artificial intelligence in the administrative procedure[30] could present some difficulties because of the European legal framework, which bans an automated decision-making process when rights and freedoms could be affected under Article 22 of the General Data Protection Regulation (GDPR). Some concerns arise also because of the 'black box' mechanism of artificial intelligence functioning (especially in machine learning or deep learning applications), which creates some collisions with administrative principles such as transparency, accountability, or motivation. In the following, we focus on some key decisions to highlight how Italian judges try to tackle digital revolution by making use of the principles of administrative law.

With regards to one of these decisions, number 9227 of 2018, the administrative judge stated the instrumental role of artificial intelligence, subordinated to the autonomous judgment of civil servants. In particular, the subjects to automated administrative acts complained about the absence of human intervention in the administrative procedure, with the consequence of the mere transposition of artificial intelligence output by the procedure. Consequently, the core principles of administrative law, such as the main role of the civil servant, the right to participate provided to the recipient and the right to oppose the decision as well as the obligation to provide reasoning for the administrative decision, seem not to have been respected. The administrative court declared that the high number of participants in a public competition for occupying a job (in this case, the competition referred to candidates to the status of public teacher) is not sufficient to justify the complete automation of the procedure. It is affirmed that some legal

28 See : *Commission nationale de l'informatique et des libertés*, Deliberation n. 119/2018, available at: www.legifrance.gouv.fr.

29 For an interesting analysis of this case study, see: Avanzini 2019. 126–135.

30 On the relationship between artificial intelligence and administrative procedure, see: Galetta–Corvalan 2019, Cavallaro–Smorto 2019, Viola 2018, Alberti 2019. 141–155; previously, Masucci 2011, Fantigrossi 1993.

arrangements, such as participation, transparency, and access to the administrative acts, could not be limited by the use of artificial intelligence systems.[31]

The administrative judge also affirmed that the evaluative, cognitive, and intellectual approach of human activities must not be replaced by the artificial intelligence system and, consequently, that such systems must be considered just as instruments for the administrative procedure. These instruments seem to be useful for analysing a high number of candidates or for doing some repetitive tasks. Nevertheless, these artificial intelligence instruments could not substitute the 'cognitive, intellectual and judgment activities that just a preliminary analysis made by a civil servant could do'.[32]

Additionally, the Court considered that the unsupervised use of artificial intelligence is in contradiction with the Italian administrative legal framework because it has to be considered as involving an administrative activity. In particular, articles 7, 8, 9, 10 and 10bis of the main act of administrative procedure (Act No 241, issued on 7 August 1990) represent the substance of the relationship between the public administration and recipients, based on participation rights. Article 3 of the previously mentioned Act refers to the right to reasoning for administrative acts, which could not be eliminated to preserve the right of defence of recipients as well as the power of the judge to know the logical procedure underpinning the decision adopted. This latter aspect is strictly connected to the notion of external full judicial review because knowing the steps adopted by civil servants makes it easier for the judge to determine the logic and reasonableness of an administrative decision.

Other decisions by the administrative regional courts (TAR Puglia, sez. I, no. 806 of 27 June 2016 and, more recently, TAR Lazio, sez. III, no. 8076 of 18 July 2018) recognize – even if with a different logical procedure – that '[a] digitised procedure applied to the administrative procedure must be put into a servant approach and, consequently, it is forbidden that technical biases obstruct the relationship between public administration and citizens'.[33] Moreover, in the same vein, the Council of State in Sentence No 5136 of November 2017 affirmed that any administrative rejection is legal if it is the consequence of a technical failure. In this specific case, a citizen submitted an online request, but his application was refused because the time limit had been exceeded due to a technical failure. Consequently, the Supreme Court revoked the rejection and attributed the responsibility for the delay to the service operator.

31 On the relationship between public law and ethics of algorithms, see: Casonato 2019. 101–130, Crisci 2018, Simoncini–Suweis 2019.

32 This idea is repeated in various sentences of administrative courts such as the Administrative Regional Court of Lazio, nos. 9225, 9226, 9227, and 9230 of 2018.

33 TAR Lazio, sez. III-bis, sent. no. 8312/2016, available at: https://www.giustizia-amministrativa.it/. Translation by the author.

Additionally, other decisions[34] consider that every *de facto* exclusion from a public competition because of a technical failure is illegal. Judges recognize that the algorithmic decision is qualified as a substantial administrative activity, and, consequently, it has to respect the principles of administrative procedure.

In each of the previous decisions, it is clear that judges want to protect citizens from decision-making algorithmic administrative procedures that could be illegal because of their contrast with the core principles of public administration such as transparency, accountability, impartiality, and legality of the action.[35]

Admittedly, the algorithmic decision-making process is rather difficult to unbundle, particularly when it comes to deep learning and machine learning mechanisms based on self-learning methods. These highly advanced technologies use big data extensively and, after an initial programming phase, their reasoning could not be known even to the coder.

4. Conclusions

Finally, it is possible to affirm that the introduction of artificial intelligence in the public sector presents a positive and a (quite) negative side at the same time. The development of artificial intelligence could improve the efficiency of public bodies: advanced technologies seem to better address and allocate resources as well as personalized services. Artificial intelligence and big data analysis represent a new way of running public management. The opportunities arising from the big data paradigm allow public bodies to analyse and capture deeply what happens in the local system: thanks to the Internet of Things, devices as well as predictive governmental analyses may provide better public services.

In particular, a new management paradigm arises: these advanced technology tools allow public administration to identify indicators and to quickly collect insights as well as to integrate, unify, and analyse data from different sources. This digital revolution makes it possible to shift 'from fact-free policy to rational and evidence-based rules',[36] where 'a proactive mode of operation based on mathematical models'[37] dominates.

Even if technologies are neither good nor bad, probably it is their use that could create some concerns. Each algorithm implies some policy choices as well as subjective judgements about what data to use, how to weigh the variables and data. Although algorithms are based on mathematical and statistical methods, it would

34 TAR Lazio, sez. III-bis, sent. no. 2272/2018; TAR Lazio, sez. III-bis, sent. n. 11786/2016, available at: https://www.giustizia-amministrativa.it/.

35 Similar interpretation is supported also by the *Commission nationale de l'informatique et des libertés*, Deliberation no. 119/2018, available at: www.legifrance.gouv.fr.

36 Ranchordas–Klop 2018. 12.

37 Appel et al. 2014. 172.

be misleading to think that they are completely objective. Programmers intervene deeply on the choice and the weight of data as well as on the procedural process. On top of this, the decision-making power of big tech firms, which benefit from more competence and skill than the public sector does, is growing in managing public interests.

The problem related to the programming and designing of algorithms[38] is that they 'can privilege different stakeholders in a decision'[39] with a high risk of discrimination. Probably this is the main concern that scholars have to deal with to preserve public interest: public bodies should collaborate with private vendors to help determine if decisions could have an unfair impact on individuals. The algorithms could make mistakes in the decision-making process because of incorrect or out-of-date inputs, but, at the same time, the usefulness of this decision-making process is evident. Consequently, a balancing act between the efficiency and the fairness of artificial intelligence in the public sector is necessary.

Because of the necessity to assess and correct an algorithm quickly and without obstacles, two solutions may be proposed.

On the one hand, the procurement's call for tender must have considered some provisions where civil servants and programmers should work and program together to guarantee that the algorithm pursues the public goal individualized and, at the same time, to allow public bodies to assess and control how the algorithm works.

The idea is to create a strict collaboration between public bodies and engineers through guidelines on the legal framework and technological measures for compliance, to be adhered to by programmers. The main goal is to reach a solution which is 'legal by design' and which allows for the preservation of administrative principles in the designing phase.

On the other hand, the review of mistaken algorithmic decisions should take place in the administrative procedure and not during judicial review. The proposal is that public bodies could adopt some act *ad hoc* to review and quickly change the wrong content of artificial acts before the judicial review takes place. This proposal intends to guarantee the rights and freedoms of recipients straightforwardly as well as to respect the European legal framework related to the ban of solely automated decision making. Artificial intelligence, big data analysis, and other advanced technologies should be fostered in the public sector to deliver more efficiency and effectiveness. This is possible thanks to a predictive approach, for re-thinking the managing of services for the community, for reaching the 'smart-city model'.

Conversely, the use of artificial intelligence or other technologies in the decision-making process should be carefully handled as it impacts individual rights. Previous case studies, analysed above, demonstrate that algorithms are

38 On the transparency of algorithm, see: Brauneis–Goodman 2018.

39 Diakopoulos 2015. 400.

value-laden, even if they are programmed with a statistical method. Some biases might have been caused by the 'tuning' of the algorithms designed by a private programmer as well as the entering of uncorrected data. Digital revolution forces the reflection on private–public partnership because of the increasing necessity to constantly assess, check, and eventually correct the algorithm used.

At the same time, the judicial review of mistaken algorithmic decisions seems to appear too cumbersome and ineffective. Thus, restoration of potential harm should be anticipated in an administrative procedure where the act could still be corrected.

Furthermore, the role of human intervention should not be underestimated even in the automated decision-making process not only because of the GDPR but, above all, because of the idea that the implementation of administrative law principles should not be utterly replaced by machines.

Nevertheless, instead of adopting artificial intelligence, it is preferable to adopt a 'human-in-the-loop' approach according to the idea that humans should oversee the automated process and, if it is necessary, should intervene and control it. Ultimately, the automated decision-making process should be admitted, with the proper guarantees, if there is a right to have a human decision.

References

ALBERTI, I. 2019. Artificial Intelligence in the Public Sector: Opportunities and Challenges. *Eurojus* (special issue): Big Data and Public Law: New Challenges beyond Data Protection: 149–163.

APPEL, S. et al. 2014. Predictive Analytics Can Facilitate Proactive Property Vacancy. *Technological Forecasting and Social Change* 89: 161–173.

AVANZINI, G. 2019. *Decisioni amministrative e algoritmi informatici. Predeterminazione analisi predittiva e nuove forme di intellegibilità*. Napoli.

BABUTA, A. 2017. Big Data and Policing. An Assessment of Law Enforcement Requirements, Expectations and Priorities. *Royal United Services Institute for Defence and Security Studies*, occasional paper 2017: entire issue.

BONFANTI, A. 2018. Big data e polizia predittiva: riflessioni in tema di protezione del diritto alla privacy e dei dati personali. *Rivista di diritto dei media* 3: 206–218.

BORGOGNO, O.–COLANGELO, G. 2019. Data Sharing and Interoperability: Fostering Innovation and Competition through APIs. *Computer Law & Security Review* 5: https://www.sciencedirect.com/science/article/pii/S0267364918304503.

BRAUNEIS, R.–GOODMAN, E. P. 2018. Algorithmic Transparency for the Smart City. *20 Yale J.L. & Tech*: https://digitalcommons.law.yale.edu/yjolt/vol20/iss1/3.

CASONATO, C. 2019. Intelligenza artificiale e diritto costituzionale: prime considerazioni. *Diritto pubblico comparato ed europeo* (special edition): 101–130.

CARRER, S. 2019. Se l'amicus curiae è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin. www.giurisprudenzapenale.com.

CAVALLARO, M. C.–SMORTO, G. 2019. Decisione pubblica e responsabilità dell'amministrazione nella società dell'algoritmo. *Federalismi.it*, 16: https://www.federalismi.it/nv14/articolo-documento.cfm?Artid=40182&content=Decisione%2Bpubblica%2Be%2Bresponsabilit%C3%A0%2Bdell%E2%80%99amministrazione%2Bnella%2Bsociet%C3%A0%2Bdell%27algoritmo&content_author=%3Cb%3EMaria%2BCristina%2BCavallaro%2Be%2BGuido%2BSmorto%3C%2Fb%3E.

CRISCI, S. 2018. Intelligenza artificiale ed etica dell'algoritmo. *Foro Amministrativo* 10: 1787–1810.

DALY, P. 2019. *Artificial Administration in Action: The Robo-Debt Scandal*. https://www.administrativelawmatters.com/blog/2019/04/11/artificial-administration-in-action-the-robo-debt-scandal/.

DI PORTO, F.–RANGONE, N. 2013. Cognitive-Based Regulation: New Challenges for Regulators? *Federalismi. Rivista di dirittopubblicoitaliano, comunitario e comparato* 20: https://ssrn.com/abstract=2382470.

DIAKOPOULOS, N. 2015. Algorithmic Accountability. *Digital Journalism* 3: 398–415.

FANTIGROSSI, U. 1993. *Automazione e pubblica amministrazione*. Bologna.

GALETTA, D. U.–CORVALAN J. G. 2019. Intelligenza artificiale per una Pubblica amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto. *Federalismi.it*: https://www.federalismi.it/nv14/articolo-documento.cfm?Artid=38014.

KITCHIN, R.–LAURIAULT, T. P.–MCARDLE, G. 2015. *Knowing and Governing Cities through Urban Indicators, City Benchmarking and Real-Time Dashboards*. *Regional Studies, Regional Science* 2(1): 6–28.

MACIEJEWSKY, M. 2017. To Do More, Better, Faster and More Cheaply: Using Big Data in Public Administration. *International Review of Administrative Sciences* 83: http://journals.sagepub.com/doi/metrics/10.1177/0020852316640058.

MASHAW, J. 1983. *Bureaucratic Justice: Managing Social Security Disability Claims*. New Haven.

MASUCCI, A. 2011. *Procedimento amministrativo e nuove tecnologie. Il procedimento amministrativo elettronico ad istanza di parte*. Turin.

OSWALD, M.–GRACE, J.–URWIN, S.–BARNES, G. C. 2018. Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality. *Information & Communication Technology Law* 27: 223–250.

PINOTTI, G. 2019. Automated Administrative Procedure and Right of Access to Source Code. *Eurojus* (special issue): Big Data and Public Law: New Challenges beyond Data Protection: 126–133.

RANCHORDAS, S.–KLOP, A. 2018. Data-Driven Regulation and Governance in Smart Cities. In: *Research Handbook on Data Science and Law*. Cheltenham (UK)–Northhampton (USA).

SCASSA, T. 2017. Law Enforcement in the Age of Big Data and Surveillance Intermediaries: Transparency Challenges. *Scripted* 14: 239–284.

SIMONCINI, A.–SUWEIS, S. 2019. Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale. *Rivista di filosofia del diritto* 1: 87–106.

TRESCA, M. 2018. I primi passi verso l'Intelligenza artificiale al servizio del cittadino: brevi note sul libro bianco dell'Agenzia per l'Italia digitale. *Medialaws* 3: entire issue.

VIOLA, L. 2018. L'intelligenza artificiale nel procedimento e nel processo amministrativo: lo stato dell'arte. *Federalisimi.it* 21:2–44.

WRIGHT, B. 2014. Technology and the Rule of Law in the Digital Age. *Notre Dame Journal of Law, Ethics & Public Policy* 2: 705–710.

Online Sources

AI-Augmented Government. Using Cognitive Technologies to Redesign Public Sector Work. https://www2.deloitte.com/content/dam/insights/us/articles/3832_AI-augmented-government/DUP_AI-augmented-government.pdf.

Commission nationale de l'informatique et des libertés. *Deliberation n. 119/2018*, www.legifrance.gouv.fr.

Senate Standing Committee on Community Affairs (Australia). *Design, Scope, Cost-Benefit Analysis, Contracts Awarded, and Implementation Associated with the Better Management of the Social Welfare System Initiative*: https://www.aph.gov.au/parliamentary_business/committees/senate/community_affairs/socialwelfaresystem.

The Rule of Law on the Internet and in the Wider Digital World. https://rm.coe.int/16806da51c.

http://legacy.fbk.eu/it/news/fbk-collabora-con-il-miur-labuonascuola.

https://binghamcentre.biicl.org/categories/digital-age.

https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

https://www.uscis.gov/emma.

Acta Univ. Sapientiae, Legal Studies, **8**, 2 (2019) 167–177

DOI: 10.47745/AUSLEG.2019.8.2.02

La France contre les robots – At the Crossroads of Humanity, Law, and Prophetic Utterance

Simona Catrinel Avarvarei
Senior Lecturer
Ion Ionescu de la Brad University of Iaşi, Iaşi
E-mail: catrinel_04@yahoo.co.uk

Nicoleta Rodica Dominte
Senior Lecturer
Alexandru Ioan Cuza University of Iaşi, Iaşi
E-mail: nicoleta_dominte@yahoo.com

Abstract. 'On ne comprend absolument rien à la civilisation moderne si l'on n'admet pas d'abord qu'elle est une conspiration contre toute forme de vie intérieure', wrote French journalist and novelist Georges Bernanos (1888–1948) towards the end of the Second World War and his self-imposed exile in Brazil, in his last completed volume of essays, *La France contre les robots*, published in 1947. More than half a century stands between the nib of the author's quill and the modern reader, leaving the text, its effervescent polemic and abysmal, rhetorical depths uncorroded and infinitely topical. The hermeneutics this article steps into was as complex at the time the essay was written as it is now, concerned as it is with the relationship between man and machine. Aware that mechanization has already started to (re)write history as we know it, Georges Bernanos is most concerned with the fact that 'la civilisation des machines est celle de la quantité opposée à celle de la qualité' in a paradigm which encourages 'd'une manière presque inimaginable l'esprit de cupidité' and whose most dramatic effect 'n'est pas dans la multiplication des machines, mais dans le nombre sans cesse croissant d'hommes habitués, dès leur enfance, à ne désirer que ce que les machines peuvent donner'. With studies of law and literature and a profound understanding of the falls and decays of the human soul, treasuring that 'supplément d'âme' Henri Bergson speaks about, Bernanos has constantly sought to explore the perilous trails of self-estrangement mechanization, this 'modern era', as it is often referred to, opens in a myriad of facets and reflexions that urged him say that 'nous n'assistons pas à la fin naturelle d'une grande civilisation humaine, mais à la naissance d'une civilisation inhumaine qui ne saurait s'établir que grâce à une vaste, à une immense, à une universelle stérilisation des hautes valeurs de la vie'. What he tries to defend is the uniqueness and singularity of man, his complexity, and not

to demonize machines and their part in reconfiguring progress, in any and all of its aspects. Danger, he warns, 'n'est pas dans la multiplication des machines, mais dans le nombre sans cesse croissant d'hommes habitués, dès leur enfance, à ne désirer que ce que les machines peuvent donner'. The key in which we intend to approach Georges Bernanos's *La France contre les robots* plays with the dichotomy of the 'productive man', epitome of the technical society, an offspring of the Anglo-Saxon skill and labour doctrine, more mechanical in its philosophy than the French ideological legal scheme of interest in the 'impact of a personality in a work of the spirit/mind'[1] of the 'contemplative man'. Whilst the first is merely a reflexion of his age, estranged from his own self, though very much a master of his time, the latter becomes the depository of the writer's hopes and symbol of creative humanity.

Keywords: human being, machine, self, law's governance, artificial intelligence

'Literature always anticipates life.'
Oscar Wilde

1. Introduction

Ezra Pound believed that 'Great literature is simply language charged with meaning to the utmost possible degree', thus acknowledging the power of creative and momentous narrative not only to weave the story of the present but also to envisage and mould the map of the future. 'Literature adds to reality, it does not simply describe it', would continue C. S. Lewis, firmly believing that fiction is not necessarily fictitious as it is thought-provoking and trailblazing.

Born in Paris on 20 February 1888, George Bernanos held a double degree in law and letters from the University of Paris. After a meandering journey through a most tumultuous early twentieth century, the horrors of the Great War inflicting physical injuries on his body, as he served as a corporal in the French Cavalry and received a chest wound, Bernanos turned to writing as 'a means of escape from this disgusting era'.[2] Following the publication of his first novel, *Sous le soleil de Satan*, in 1926, George Bernanos became one of the most important writers of French literature, as Léon Daudet prophesied it would happen in an article published in *Ecrivains et artistes* on April 7:

> Demain le premier livre, le premier roman d'un jeune écrivain, M. Georges Bernanos auteur de *Sous le soleil de Satan*, sera célèbre. Je dirai de lui, comme je le disais naguère de Marcel Proust – hélas ! - qu'une grande force,

1 Kearns 2013.
2 Bernanos 1945.

> intellectuelle et imaginative, apparaît au firmament des lettres françaises. Mais cette fois synthétique, et non plus analytique, et dans un genre à ma connaissance encore inexploré et qui est le domaine de la vie spirituelle, des choses et des corps commandés par les âmes.[3]

Nevertheless, Bernanos failed to consider himself a genuine author, as he acknowledged in *A Diary of My Times*:

> I am no author. The sight alone of a blank sheet wearies my spirit, and the sheer physical isolation imposed by such work is so distasteful to me that I avoid it as much as I can. […] I write at café tables because I cannot long be deprived of the human face and voice, which I have tried to render with dignity. Let clever folk suppose that I sit 'observing' my fellow men. I observe nothing. Observation never leads to much… I scribble in cafés just as I used to scribble in railway carriages, in order not to be taken in by figments of my own imagination, in order at a glance to re-discover, in the unknown person opposite, my own fair measure of joy or sorrow. No–I'm not an 'author'. Had I been a real one, I never should have waited until I was forty before I published my first book. […] A vocation is always a call to action–*vocatus*–and every call must be passed on. Those to whom I call are obviously few. They will alter in nothing the ways of the world. Yet it is for them–for them that I was born.

Bernanos's legacy has spanned over the past century and proved its stamina and perceptiveness in the current whirlwind of ideas concerning one of humanity's greatest challenges – superhuman artificial intelligence. His writing most certainly did not concern only a few of his contemporaries, as the French polemicist would consider, for his voice came to speak the language of many more people, some of whom forge the history of AI as we write (speak).

2. Beyond Literature's Path, from Story to Science in the mid-1950s

Far beyond that path one may notice, looking through a telescope, Antoine de Saint-Exupery's legendary asteroid B-612 on which dwells, alongside the Little Prince and his Rose, the fervid and resolute French spirit. Most famous is the scene in which the pilot, who has crash-landed in a desert, encounters a small boy; the latter asks him to draw him a sheep, and the narrator obliges. The task is by no

3 El Gammal 2012.

means easy as he seems to lack the skills of drawing a sheep, any sheep for that matter, let alone the one the Little Prince, for this is how the narrator would refer to the child, dreams of – 'one that will live for a long time'. The artistically challenged castaway only managed to hide one inside a box, much to the awe and joy of the Little Prince, who, given the opportunity, one can only speculate, might have turned to artificial intelligence (AI) for the perfect anti-age remedy. Sadly, though, when the Little Prince wanted to tell his story to the whole world, he realized that it was too old and not at all willing to listen to new voices and stories about the human soul and some of its most complex personal relationships.

2.1. The Fable…

One such voice was George Bernanos, who, like T. S. Eliot, Earnest Hemingway, James Joyce, William Faulkner, or George Orwell, depicts a disheartening perspective upon life devoid of any genuine human vibration and enduring spiritual values. His essay *La France contre les Robots*, published four years after the famous novella, in 1947, and anonymously translated into English as *Tradition of Freedom* (1950), focussed on the mechanization of human life and the dramatic dwindling of inner godly expression. He deplores the fall-off and simplification of man from the privileged status of bearer of divine light to a banal performer and mere doer of things.

> Dans la lutte plus ou moins sournoise contre la vie intérieure, la Civilisation des Machines ne s'inspire, directement du moins, d'aucun plan idéologique, elle défend son principe essentiel, qui est celui de la primauté de l'action.[4]

The new actor on the stage of this world is *homo faber*, a threat to himself and the world he so frenziedly forges:

> Human civilization, I've said it, is the whole man, the brain, the heart and the viscera, soul and body. Here is before us the man left to the mercy of his own hands, his rebellious hands, his hands suddenly multiplied by technique and mechanics, the man attacked by his hands, stripped by them, left naked like a worm who expects to be dismembered little by little, piece by piece, fibre with fibre, into total disintegration. For the atomic bomb, do not be deceived in this regard, is still a hand, though so fine, so subtle, that it breaks down the atoms as one breaks the pea-berries out of a pod. Here the technique, the science of hands, is caught in a flagrante, like the agile hand of a thief in the pocket of a looky-loo. For it is no longer about dominating matter, it is about its destruction.[5]

4 Bernanos 1947.
5 Bernanos 1953.

Like Karel Capek, the equally famous Czech writer who had not only coined the world *robot* and introduced it to the whole world in 1921 with the premiere of his three-act play *Rossum's Universal Robots* but who equally and powerfully deplored the dehumanization of man through technology, Bernanos was genuinely concerned about man's ability to fully understand and foresee all the implications of such advances. Norbert Wiener, the American mathematician and philosopher considered to be the father of cybernetics, asserts that as long as automata can be built, either in blueprints or effectively, it will only echo the natural inquisitiveness of the human mind; nevertheless, he warns us some of the reasons that lead to robotization (might) cross the line of legitimate curiosity.[6] The strokes are bleeding on the canvas of his thoughts, flooded by the apocalyptical scenario of hopelessness, emptiness, and loss of human self and godly values. Destruction and annihilation seem to become the exclusively foreseen scenario of a civilization that fragments itself and slips into trivial nonexistence. The words compose the texture of the writer's soul, desperate to (re)create balance and harmony and to instil peace and faith in the downheartedly colourless landscape of some mechanic reality that tends to become the only framework of a new entropic universe.

> Les âmes! On rougit presque d'écrire aujourd'hui ce mot sacré. Les mêmes prêtres imposteurs diront qu'aucune force au monde ne saurait avoir raison des âmes. Je ne prétends pas que la Machine à bourrer les crânes est capable de débourrer les âmes, ou de vider un homme de son âme, comme une cuisinière vide un lapin. Je crois seulement qu'un homme peut très bien garder une âme et ne pas la sentir, n'en être nullement incommodé ; cela se voit, hélas ! tous les jours. L'homme n'a de contact avec son âme que par la vie intérieure, et dans la Civilisation des Machines la vie intérieure prend peu à peu un caractère anormal.[7]

It is against such a projection that Bernanos directs his tirade and serious concern for the decline of human ideals and effacement of human sensitivity. *La France contre les robots* is about the philosopher's anguish regarding the articulation of a 'brave' new world with shallow appearances to defend and materialistic goals to construct. 'Aller vite? Mais aller où?' – asks the writer, urged as he is by the need to understand the race of the modern human actor, more concerned with his ephemeral material legacy rather than with his eternal spiritual legacy, and to make people understand that the future is not as much a projection of one's pragmatic becoming as it is a quest of a deeper spiritual nature. Swooshing right by Life and its quintessential core instead of breathing in the complex concert it entangles becomes the new and most shattering *modus vivendi*. He is not against technological progress, which he admits and supports, as the threat is not posed by

6 Wiener 1964.
7 Bernanos 1947.

the 'multiplication des machines' as it lies with the 'nombre sans cesse croissant d'hommes habitués, dès leur enfance, à ne désirer que ce que les machines peuvent donner'.[8] His battle is not against technology but against the conscience- and soul-annihilating impact of scientific progress upon ordinary people, sometimes much too eager to embrace all technical novelty without demur.

> Mais à quoi bon vous dire quel type d'homme elle prépare. Imbéciles ! n'êtes-vous pas les fils ou les petits-fils d'autres imbéciles qui, au temps de ma jeunesse, face à ce colossal Bazar que fut la prétendue Exposition Universelle de 1900, s'attendrissaient sur la noble émulation des concurrences commerciales, sur les luttes pacifiques de l'Industrie ?… À quoi bon, puisque l'expérience de 1914 ne vous a pas suffi ? Celle de 1940 ne vous servira d'ailleurs pas davantage. Oh ! ce n'est pas pour vous, non ce n'est pas pour vous que je parle ! Trente, soixante, cent millions de morts ne vous détourneraient pas de votre idée fixe : « Aller plus vite, par n'importe quel moyen. » Aller vite ? Mais aller où ? Comme cela vous importe peu, imbéciles ! Dans le moment même où vous lisez ces deux mots : Aller vite, j'ai beau vous traiter d'imbéciles, vous ne me suivez plus. Déjà votre regard vacille, prend l'expression vague et têtue de l'enfant vicieux pressé de retourner à sa rêverie solitaire… « Le café au lait à Paris, l'apéritif à Chandernagor et le dîner à San Francisco », vous vous rendez compte !… Oh ! dans la prochaine inévitable guerre, les tanks lance-flammes pourront cracher leur jet à deux mille mètres au lieu de cinquante, le visage de vos fils bouillir instantanément et leurs yeux sauter hors de l'orbite, chiens que vous êtes ! La paix venue vous recommencerez à vous féliciter du progrès mécanique. « Paris-Marseille en un quart d'heure, c'est formidable ! » Car vos fils et vos filles peuvent crever : le grand problème à résoudre sera toujours de transporter vos viandes à la vitesse de l'éclair. Que fuyez-vous donc ainsi, imbéciles ? Hélas ! c'est vous que vous fuyez, vous-mêmes – chacun de vous se fuit soi-même, comme s'il espérait courir assez vite pour sortir enfin de sa gaine de peau… On ne comprend absolument rien à la civilisation moderne si l'on n'admet pas d'abord qu'elle est une conspiration universelle contre toute espèce de vie intérieure. Hélas ! la liberté n'est pourtant qu'en vous, imbéciles ![9]

George Bernanos deliberately changes the register and chooses the companionship of punitive words only to cast the 'evil' spell away. Timon's of Athens 'fools of fortune' (Act III, scene VI) are just as much his words as they are William Shakespeare's.

> Imbéciles de droite et de gauche, chiens que vous êtes, si vous vous grattez si furieusement, c'est que vous vous sentez, au fond, tous d'accord, vous

8 Bernanos 1947.
9 Bernanos 1947.

> savez tous très bien qu'à la Civilisation des Machines doit logiquement correspondre la guerre des machines. Assez de grimaces, hypocrites ![10]

Beyond these bleak projections, spanning across the entire spectrum of the social life, there is an immense suffering and a hesitant world, insufficiently willing to stop from its pace and listen.

> Technology is alchemy; it is the self-fulfilment of nature in place of the self-fulfilment of the life that we are. It is barbarism, the new barbarism of our time, in place of culture. Inasmuch as it puts the prescriptions and regulations of life out of play, it is not simply barbarism in its most extreme and inhumane form that has ever been known—it is sheer madness.[11]

These are the thoughts of another French philosopher, Michel Henry, published forty years later, in 1987; the same articulate concern for the preservation of life and its intrinsic values now at risk more than ever with the development of science and technology at the expense of humanitarianism, art, ethics, emotion, and religion.

> …le progrès n'est plus dans l'homme, il est dans la technique, dans le perfectionnement des méthodes capables de permettre une utilisation chaque jour plus efficace du matériel humain.'[12] Ironically, the alchemy both philosophers refer to seems to be a reversed one since all it does, in the end, is to turn gold into lead. Bernanos operates with clinical precision and diagnoses a most troublesome anamnesis, which he quintessentially describes as 'décoloration de la conscience – la maladie des consciences pales.[13]

2.2. …and the Science

The moment *La France contre les robots* was published, the frenzy fever of scientific research that was constantly forcing the frontiers of knowledge agglutinated around such telling theories, like the one predicated by Dennis Gabor, the inventor of holography and Nobel-prize laureate for physics in 1971 – 'all that can be accomplished from a technical perspective must be accomplished, regardless of the ethical costs implied.'[14] After the Second World War, the scientific community was firm in its belief that any time soon artificial intelligence would articulate a robust sense of self-awareness just as Norbert Wiener understood that 'The world

10 Bernanos 1947.
11 Henry 2012. 52.
12 Bernanos 1970 [1947].
13 Bernanos 1970 [1947].
14 Alexandre–Besnier 2019. 15.

of the future will be an even more demanding struggle against the limitations of our intelligence, not a comfortable hammock in which we can lie down to be waited upon by our robot slaves.'[15] Once the threshold of the 21st century was crossed, the ideas have remained unaltered, as the fears still linger, and we only have to listen to Laurent Alexandre, a most reputed French urological surgeon, author, entrepreneur, expert on transhumanism, and Head of NBIC Finance to understand that not much has changed – 'the fusion between human and machine will mean the annihilation of the biological man'.[16] Irvin John Good, British mathematician, who worked as a chief statistician at Bletchley Park with Alan Turing and who continued to work with Turing on the design of computers after the Second World War, would define AI with the following words:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion', and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.[17]

Although the pioneers of artificial intelligence 'did not contemplate the possibility of greater-than-human AI',[18] Alan Turing wholeheartedly believed in the existence of machine intelligence capable of constantly 'improving its own architecture'.[19] The robot Alan Turing was planning to design would not be an augmented humanoid but a *brain* which can be trained and taught to think, and in this respect he seemed to share a different opinion from that of his former philosophy professor at Cambridge, Ludwig Wittgenstein, who seeded his reluctance on the possibility of a machine *to think* precisely in the verb itself.[20]

Years later, after master and disciple would have embarked upon their celestial journey, in the summer of 1956 at Dartmouth College, in the United Kingdom, a Summer Project was initiated in an attempt to 'find how to make machines that use language, form abstractions and concepts, solve kinds of problems now reserved for humans'.[21] It was clear that artificial intelligence and its journey was no longer a question of vague possibility but a scientific promise and certainty which would

15 Wiener 2019 [1964].
16 Alexandre–Besnier 2019. 47.
17 Good 1965. 33.
18 Bostrom 2017. 5.
19 Bostrom 2017. 34.
20 Boyle 2014. 103.
21 Bostrom 2017. 6.

only expand with its travel through time and cultures. The costs, nevertheless, were unforeseeable. Bernanos held that 'Un monde gagné pour la Technique est perdu pour la Liberté'[22] and Freedom is the very scaffolding of every spiritual evolution, evermore so of the French spirit 'Il y a une tradition française de la Liberté. En 1789, tous les Français, pour un moment du moins, ont communié dans cette tradition, chacun selon l'étendue de ses connaissances ou la force de son esprit, mais avec une foi simple, unanime.'[23]

In June 1949, Sir Geoffrey Jefferson, professor of neurosurgery at the University of Manchester, made the following statement in the Lister Speech entitled 'The Mind of Mechanical Man':

> Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain – that is, not only write it but know that it had written it. No machine could feel pleasure at its success, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or miserable when it cannot get what it wants.[24]

But then few people believed that man would eventually fly to the Moon and back!

George Bernanos must have been right as nowadays Elon Musk, the founder of Paypal, Hyperloop, SolarCity, Tesla, and SpaceX, warns us that 'AI can turn into something far more dangerous than the nuclear weapons.'[25] The gloomy perspective is that in order to be able to measure up to AI-endowed robots, some, and not few, scientists believe that we have to hybridize ourselves with AI just to avoid a feeling of inferiority. In an open letter published on 27 July 2015 and signed by more than one thousand renowned scholars, among whom Elon Musk (businessman), Noam Chomsky (linguist), Stephen Hawking (astrophysicist), and Bill Gates (founder of Microsoft), advocated that the AI will pose serious problems to humanity; and just a few months later Hawking would write that the 'development of a totally complete AI may mean the end of the human race'.[26] To all that, we could add Ray Kurzweil's prophecy that by 2045 a non-biological form of intelligence will surpass our own, and there is the even more serious risk that this AI will destroy what is human in us, depriving us of the will and power to decide our own fate. In fact, ever since the invention of the steam engine and the change brought about by the Industrial Revolution, the machine had become accountable for our ever-growing sense of helplessness. It is one of the factors

22 Bernanos 1947.
23 Bernanos 1947.
24 Cf. Boyle 2014. 101–102.
25 Cf. Alexandre–Besnier 2019. 46.
26 Cf. Alexandre–Besnier 2019. 83.

that are responsible for 'the promethean shame of being oneself', as the Austrian philosopher Gunter Anders asserts.[27]

Nick Bostrom, the founding Director of the Future of Humanity Institute, Oxford University, author of the bestseller titled *Superintelligence: Paths, Dangers, Strategies*, argues that true artificial intelligence might pose a threat to humanity and its evolution far greater than any other previous technological breakthrough. 'This is quite possibly the most important and most daunting challenge humanity has ever faced. And – whether we succeed or fail – it is probably the last challenge we will ever face.'[28] 'Before the prospect of an intelligence explosion, we humans are like small children playing with a bomb', he concludes. 'We have little idea when the detonation will happen.' The scholar claims that there is room for only one intelligent species in each corner of the Universe, while 'predictions about the future development of artificial intelligence are 'as confident as they are diverse'.[29]

Who could say that there is a space of more than sixty years between the two philosophical papers and two authors of so different an intellectual background? 'Chaque invention nouvelle accroît le prestige de la Force, et fait décroître celui du Droit. Dans un monde armé jusqu'aux dents, le juge de Droit International Public finit par devenir une espèce de personnage cocasse, le survivant d'une époque disparue.'[30]

3. Conclusions

When the fire engulfed the roof and spire of the Notre Dame Cathedral on 15 April 2019, Paris knelt and prayed for its symbol; emotion filled the air of the blazing dusk, and the murmur of thoughts instilled with hope was the only utterance of millions of voices. That was the moment when France proved to the world that neurons, and not silicon and human emotions, are still at the very heart of our civilization and that silicon is merely a technological implement destined to help write the future and (hopefully) not the very future itself. George Bernanos would have been relieved to see that his fellow countrymen, heirs of the superb and glorious Hellenic civilization, measure their lives against the tolls of the legendary mediaeval minster – and that is the call of *theosis* that he so much feared for.

> Obéissance et irresponsabilité, voilà les deux Mots Magiques qui ouvriront demain le Paradis de la Civilisation des Machines. La civilisation française,

27 Cf. Alexandre–Besnier 2019. 84.
28 Bostrom 2017. V.
29 Bostrom 2017. 22.
30 Bernanos 1947.

> héritière de la civilisation hellénique, a travaillé pendant des siècles pour former des hommes libres, c'est-à-dire pleinement responsables de leurs actes: la France refuse d'entrer dans le Paradis des Robots.[31]

On Easter Eve, France refused to be anything but a country of faith, hope, and sensitiveness.

References

ALEXANDRE, L.–BESNIER, J. M. 2019 [2016]. *Pot face roboții dragoste?* Bucharest: Humanitas.

BERNANOS, G. 1945. Autobiographie. In: *Essais et écrits de combat*, *II*. Paris: Gallimard.

1970 [1947]. *La France contre les robots*. Plon.

1995 [1953]. *La liberté pour quoi faire?* Gallimard.

BOSTROM, N. 2017 [2014]. *Superintelligence: Paths, Dangers, Strategies.* Oxford: Oxford University Press.

BOYLE. D. 2014. *Codul Enigma. Alan Turing şi înfrângerea Germaniei naziste*. Bucharest: Corint.

EL GAMMAL, J. 2012. Léon Daudet critique: histoire, littérature et politique. In: Michel Leymarie–Olivier Dard–Jeanyves Guérin (eds), *Maurrassisme et littérature: L'Action française. Culture, société, politique IV*. Presses Universitaires du Septentrion. 55–69.

GOOD, I. J. 1965. Speculations Concerning the First Ultraintelligent Machine. In: Frantz L. Alt–Morris Rubinoff (eds), *Advances in Computers* 6: 31–88. New York: Academic Press.

HENRY, M. 2012. *Barbarism*. London: Continuum.

KEARNS, P. 2013. *Freedom of Artistic Expression: Essays on Culture and Legal Censure*. Hart Publishing.

WIENER, N. 2019 [1964]. *Dumnezeu şi Golemul. Comentariu asupra câtorva probleme în care cibernetica intră în contradicție cu religia*. Bucharest: Humanitas.

31 Bernanos 1947.

Acta Univ. Sapientiae, Legal Studies, **8**, 2 (2019) 179–189

DOI: 10.47745/AUSLEG.2019.8.2.03

Artificial Intelligence as a Tool to Improve the Administration of Justice?

Maria Dymitruk
Faculty of Law, Administration and Economics
University of Wroclaw, Wroclaw
E-mail: maria.dymitruk@uwr.edu.pl

Abstract. Recently, technological development made a significant impact on the administration of justice. Lawyers, both legal practitioners and academics, can no longer afford to ignore the potential that the technology offers. The development of new fields in legal informatics, such as the applicability of Artificial Intelligence (AI) in law, opened up new opportunities which have hitherto been unthinkable. In the not too distant future, lawyers will need to answer the question whether AI can be engaged in the process of judicial decision making. On the other hand, the creation of a well-functioning artificial intelligence system which can carry out numerous adjudicating activities and reasoning processes is not the only requirement for using artificial intelligence in the automation process of judicial activities. Detailed analysis of its legal compliance is needed as well. This paper analyses the admissibility of using artificial intelligence tools in the judiciary and contains considerations on ethical aspects of AI application in judicial proceedings (whether an AI system is capable of taking over the role of a decision maker in judicial proceedings, thereby replacing, or supporting the judge). The research presented in the paper may provide an impulse to start a large-scale scientific discussion on the possibility and admissibility of AI application in the judicial system and may also be the basis for formulating proposals addressed to lawmakers and policymakers.

Keywords: artificial intelligence, judiciary, court proceedings, e-court

1. Introduction

The role of science is to gaze into the future, to anticipate the possibility of a particular phenomenon's occurrence, and sometimes even to adjust reality to human needs. For many years, the goal of scientists dealing with legal informatics and computerization of the judiciary has been to adjust the law to a constantly changing technological landscape and to create legal solutions

that meet the needs of modern society. For that purpose, there were numerous attempts to use computers, electronic devices, and other modern technologies as tools for facilitating the work of lawyers: starting with bringing the electronic payment order proceedings[1] into force, through providing online access to court judgements or computerization of public registers, and ending with the introduction of the electronic court report and e-filing systems before courts. The digitization of legal information and the creation of technology supporting the preparation of legal documents played a significant role in the development of computerization. Moreover, it is worth mentioning that the automation of simple and repeatable actions to eliminate unnecessary human labour has always been one of the goals of computerization.[2] But this automation did not interfere with the process of applying the law – the core element of every judicial proceeding and the element restricted only for human beings until now.

Taking the above into account, further developments of computerization in the field of judicial proceedings are worth considering. Constant development of artificial intelligence instruments allows improving the functioning of the administration of justice. One of the ideas for such improvement is the attempt to automate judicial proceedings by creating artificial intelligence systems with the ability to decide legal cases unassisted or supported by a human judge.

2. Artificial Intelligence

There is no widely accepted definition of *artificial intelligence*.[3] It is not the purpose of this paper to present every possible meaning of this term. Our aim is to analyse the admissibility of using current artificial intelligence tools in the judiciary. To achieve it, it is enough to indicate that 'artificial intelligence' consists of various automated problem-solving techniques in cases when these problems cannot be resolved with the use of simple algorithms. The main purpose of our research on artificial intelligence is – of course – to create a system equipped with the ability of independent thinking: perception, understanding, prediction, or drawing conclusions. Speaking of artificial intelligence, creators assume that the development of artificial minds with an intelligence equal to our own or even superior to ours will eventually take place. This objective has yet to be achieved. Nevertheless, the creators of artificial intelligence methods have reached many intermediate goals. Most of them can be used during judicial proceedings. For

1 E.g. in Poland the electronic payment order procedure was introduced to The Civil Procedure Code in the Act of 9 January 2009 on the Amendment to the Civil Procedure Code and other Acts (as published in the Official Journal in 2009, number 26, item 156); http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20090260156 (accessed: 20.08.2019).

2 Gołaczyński 2010. 4.

3 See more: Russell–Norvig 2010. 1–2.

this reason, the paper deals only with 'specialized AI', i.e. artificial intelligence methods optimized around one specific task (opposite to 'general AI',[4] which is still considered to be in the future if it is attainable at all). Therefore, the 'artificial intelligence' referred to in the title of this paper shall be understood as any existing AI methods (procedures, applications, implementations) able to conduct the legal reasoning required to make a judgment in judicial proceedings. It includes but is not limited to symbolic approaches and sub-symbolic methods such as neural networks.

Due to the above, in the paper, only current achievements in the field of AI are analysed. As a result, the paper does not cover considerations on an autonomous AI judge which could be created in the future (a machine that could successfully perform any intellectual task that a human being – a human judge – can perform or a machine that is capable of experiencing consciousness). Despite this, one of the goals of the paper is to convince the reader that the application of AI in the judiciary does not have a futurological nature.

3. Research on AI & Law and Implementation of AI in the Legal Sphere

Successes of the creators of artificial intelligence have always stimulated the imagination of scientists, including lawyers. Research on relations between artificial intelligence and law has been the subject of scientists' interest since at least the 1970s.[5] For the first thirty years, science was interested mostly in knowledge-based AI systems. In the 1980s, the research was directed primarily at information extraction and information retrieval as well as the construction of so-called *expert systems* of various kinds. In the late 1980s and early 1990s, the emphasis was also placed on various logical formalisms (in particular deontic logics). Machine learning techniques began to be studied in the AI & Law community in the mid-2000s, and the data analytics started to be taken seriously in the early 2010s.[6]

In the beginning, all initiatives in the field of AI & Law were purely academic, but over time businesses took an interest in AI tools in legal practice. And as a result now, for several years, we have been dealing with a legal tech boom. In a legal sphere, AI systems are most frequently applied in advanced case-law search

4 Artificial general intelligence (AGI) refers to systems that exhibit intelligence comparable to the human one. Machines equipped with general AI have the capacity to understand or learn any intellectual task that a human being can.

5 Actually, papers on preliminary logic-based AI can be traced back to the early 1950s, but the phrase AI & Law started to be used in the 1970s.

6 Coenen–Bench-Capon 2017.

engines as assistance in drafting legal documents, in predictive analytics systems, as automated verification of legal compliance, or as legal aid chatbots. The use of AI systems to support the work of legal practitioners has initially been observed in the private sector. Let us mention a few examples:

1) *ROSS Intelligence* in the U.S.A. It is created by the IBM legal research service for U.S. law and is powered by artificial intelligence. ROSS is based on the now famous Watson – a question-answering computer system capable of answering questions posed in natural language, developed in IBM's DeepQA project.[7] Watson is well-known for winning the quiz show Jeopardy! while competing against human champions of this show.[8]

2) *Predictice* in France. It is a predictive analytics tool for estimating a success rate of court proceedings. Authors of Predictice claim that the system can analyse one million judicial decisions within 1 second, and in the last two years they started cooperation with over four hundred lawyers.[9]

3) *Luminance* in the UK. It is a machine learning system which improves legal analytics by reading, understanding, and learning from analysed documents. Luminance was launched in 2016, and since then it has been said to be used by over 14 of the global TOP 100 law firms.[10] Its pattern recognition technology, advanced statistical probability analysis, supervised and unsupervised machine learning methods are said to allow identifying similarities, differences, and anomalies at all levels of the review of legal documents; thus, the system can be used in, e.g., due diligence or compliance analysis.

Recently, the possibilities offered by the AI systems have been attracting increasing attention from governments and public authorities. As an example, a Brazilian project-in-progress at the Brazilian Supreme Court, called VICTOR, which was developed in partnership with the University of Brasília, aims to support the Brazilian Supreme Court by providing analysis of the cases that reach the Court, using document analysis and natural language processing tools.[11] In Europe, Latvia is exploring the possibilities for the use of the machine learning systems in the administration of justice.[12] Also, the Estonian Ministry of Justice designed a 'robot judge' that can adjudicate small claims disputes of less than €7,000. Officials hope the system can clear a backlog of cases for judges and court clerks.[13]

7 https://rossintelligence.com/ (accessed: 20.08.2019).
8 https://www.youtube.com/watch?v=P18EdAKuC1U (accessed: 20.08.2019).
9 https://predictice.com/ (accessed: 20.08.2019).
10 https://www.luminance.com/ (accessed: 20.08.2019).
11 Da Silva et al. 2018. 7.
12 Appendix I to the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment adopted by the Council of Europe European Commission for the efficiency of justice (CEPEJ) during its 31st plenary meeting, Strasbourg, 3–4 December 2018. 14.
13 https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/ (accessed: 20.08.2019).

But the public use of AI systems had varying degrees of success; some of the most known – and fairly controversial ones – include the HART (Harm Assessment Risk Tool): the AI-based technology created to help the UK police makes custodial decisions based on the recidivism risk assessment – it has been described as reinforcing existing biases. Similarly, COMPAS, the US Correctional Offender Management Profiling for Alternative Sanctions also presented this problem. This risk assessment algorithm was created and used to predict potential hotspots of violent crime and assess the risk of recidivism. In simple words, COMPAS was used to forecast which criminals are most likely to re-offend. COMPAS was highly efficient, but it ran a high risk of racial profiling and raised questions about non-discrimination. Through COMPAS, black offenders were seen almost twice as likely as white offenders to be labelled as posing a higher risk of recidivism but did not re-offend. The COMPAS software produced the opposite results with white offenders: despite their criminal history displaying a higher probability of re-offending, they were more likely to be labelled as a lower risk than black offenders.[14]

4. Polish Perspective: The Need for Change

The rapid development of AI techniques today allows us to create systems which may be able to support the judiciary (at least in some of the proceedings). The application of AI in the field of justice has the potential to revolutionize it by, inter alia: accelerating judicial proceedings, unifying case-law, widening access to court, and increasing cost-efficiency. It is, therefore, worth resenting the capabilities of the systems automating the civil proceedings (on the example of Poland).

Currently in Poland, all judicial proceedings are performed by human judges without any support of AI systems. On 4 January 2010, the electronic court (the e-court)[15] was inaugurated. The e-court considers pecuniary civil claims under an electronic payment order procedure. The claimant communicates with the e-court exclusively electronically, via the Internet, employing a system dedicated to the proceedings. The payment order (one of the types of judicial judgements in the Polish legal system) is issued by a judge or a court referent and then automatically

14 The NGO ProPublica analysed COMPAS assessments and published an investigation claiming that the algorithm was biased (https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing – accessed: 20.08.2019). The NGO Big Brother Watch in the UK criticized the HART system for 'unfair and inaccurate decisions, and a 'postcode lottery' of justice, reinforcing existing biases and inequality' (https://bigbrotherwatch.org.uk/wp-content/uploads/2018/07/Big-Brother-Watch-evidence-Policing-for-the-future-inquiry.pdf – accessed: 20.08.2019).

15 The 16th Civil Division of the Lublin Regional Court (now the 6th Civil Division of the Lublin-West Regional Court in Lublin).

served on the claimant, utilizing the electronic system, whereas the service of the lawsuit and the payment order on the defendant takes place in the traditional way (a hard copy of the payment order is delivered by post). The payment orders are issued only based on the circumstances indicated by the claimant in a statement of claim. It is important that the claimant must refer in the statement of claim to evidence in support of his arguments. However, such evidence need not be attached to the statement of claim. It means that during e-court proceedings the claimant is not required (as in traditional proceedings) to prove (with documents) the circumstances justifying their claim. Additionally, these circumstances are not verified by the judge with relation to the viewpoint of the defendant as the defendant does not participate in e-court proceedings. The defendant learns about the complaint when he is served with the payment order, together with the statement of claim. If the defendant disagrees with the arguments of the claimant expressed to the e-court in the payment order, they have the right to file a statement of opposition. The statement of opposition revokes the payment order. As a result, the case starts over from the beginning, but according to the rules of adversarial litigation – with equal participation of the claimant and the defendant. The rate at which the payment orders rendered by the e-court are opposed is between 18% and 26%.[16]

In the e-court, 8 judges, 64 court referents, and 55 external court referents (jointly 127 people)[17] cooperate in the rendering of decisions. According to the data published by the Polish Ministry of Justice, only in the first half of 2018, 1,334,284 civil cases were resolved by the e-court. Assuming an 8-hour working time of adjudicators[18] (as a rule, this is the maximum daily working time in the Polish legal system), by use of simple arithmetic, we can easily conclude that the average time for resolving a civil case in the e-court was as little as 5.67 minutes (and in 2017 the average time was 4.96 minutes). Of note, the total number of civil cases resolved by civil courts in Poland in the first half of 2018 was 6,530,208, while the average processing time of a civil case in the non-electronic writ of payment proceedings in Polish civil courts was 3.8 months.[19]

The above data shows that 20% of civil cases in Poland are currently examined in the e-court in electronic proceedings. The time of examination of a single civil case and the percentage at which the payment orders rendered are opposed prove that these cases do not require the increased activity of a judge. It seems that the electronic payment order proceedings may constitute a perfect ground for the implementation of activities in the field of AI and law. This would also make possible the transfer of the 127 judges adjudicating currently in the e-court to more

16 Data for 2010 – 2013 gathered by Łukasz Goździaszek (Gołaczyński–Mączyńska 2017. 213, 224– 228).
17 https://www.e-sad.gov.pl/Subpage.aspx?page_id=44 (accessed: 20.08.2019).
18 124 working days (992 working hours) passed from 1 January until 30 June 2018.
19 https://isws.ms.gov.pl/pl/baza-statystyczna/publikacje/download,2779,0.html (accessed: 20.08.2019).

complicated civil cases, in which they could entirely use their vast competences, their knowledge, and experience.

The analysis of statistical data leads to the conclusion that some types of civil proceedings in Poland are ready for full automation from a technological and a functional point of view. However, a question arises as to whether the binding legal framework of civil proceedings allows such automation. It turns out that questions about the admissibility of replacing a judge with a computer program are not completely meaningless and – even today – do not have a purely hypothetical aspect.

5. Machine + Human?

The information on algorithmic bias (as in the case of COMPAS and HART) can be surprising. Technology used to be regarded as neutral and impartial, and decision support systems powered by AI as a great tool to augment human judgement and reduce both conscious and unconscious biases. But from the perspective of machine learning algorithms, this opinion can be seen as being outdated. Data-driven decision making may reflect and amplify existing cultural prejudice and inequality.

The above-mentioned examples show that the use of AI in the justice system may present not only great advantages but also serious weaknesses. Of course, efficiency is a clear advantage of the use of AI in the judiciary, but it cannot overrule other aspects (such as human rights or ethical considerations). One of the ideas for surmounting the obstacles connected with the use of AI in judicial proceedings may be using AI systems not instead of human judges but in support of them (human judges possess wisdom and experience which could overcome AI's weaknesses). Taking the above into account, two possible models of AI application in the judiciary can be identified:

(1) use of AI tools to create a system which can adjudicate cases unassisted (in such cases, the system would adjudicate instead of the judge),

(2) use of AI tools to create a support system for a judge (in this model, the system would only support the judge by finding relevant facts, analysing the case-law or reviewing the legal literature, and, finally, suggesting a decision to the judge).

At first glance, most people regard the second model (humans supported by machines) as more desirable. Psychological research, however, shows that despite appearances the use of AI as a support tool can be potentially dangerous too. It might seem that this model is neutral as the decision-making process will remain in human hands. However, it turns out that using AI only in support of a judge may have the same results as the full automation of judicial proceedings. This

results from the psychological consequences of the 'persuasiveness' of decision support systems.

J. J. Dijkstra undertook a psychological experiment examining how lawyers respond to AI-generated solutions while resolving a case.[20] There were two groups of participating lawyers, both were resolving legal cases: the first group with the support of the AI system and the second one by themselves. The experiment has shown that lawyers:

— have difficulties with the assessment of the accuracy of the automatically generated advice as they focus on the argumentation presented in favour of the solution by the system, while ignoring alternative solutions;

— trust the system too much, and as a result they carelessly accept the system's advice (including incorrect solutions inserted on purpose into the experiment by the experimenters);

— in cases in which they are being advised by two entities (the system and another human), participants considered the system's advice 'to be more objective and rational than the human one' (even when the human advice was essentially identical to that provided by the system).

As a result, the participants performing legal reasoning without the support of the system achieved better results than the participants using the support system. The participants' conduct resulted from a certain psychological reaction – a desire to avoid excessive effort when processing information. The research proves that people use computer systems to evade the decision-making process and not to increase the quality of their own decisions.[21] It is therefore possible that the use of AI-generated support in the judiciary might not improve adjudication or even be detrimental to the quality of decisions rendered. Reliance on AI support systems may result in decisions regarding legal issues of citizens being made by the computer program – despite the false impression that all the guarantees supposedly provided by human adjudication are kept in place. Ignoring this fact in the legal analysis of using AI in the judiciary could bring our research and the potential application of AI in the judiciary to the level of methodological and scientific carelessness.

The presented research indicates that although there are two models for the automation of judicial proceedings (the model of replacing the human with the machine and the model of the AI system supporting the human judge), the analysis of their legal admissibility is convergent in some respect. In both cases, the effect of their functioning is similar: it is the system, not the human, who is the author of the judgment in each legal case. This circumstance was also presented in the publication with the title *Algorithms and Human Rights – Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible*

20 Dijkstra 2001. 119–128.

21 Todd–Benbasat 1994.

Regulatory Implications, prepared in March 2018 by the Committee of Experts on Internet Intermediaries (MSI-NET) of the Council of Europe:

> [g]iven the pressure of high caseloads and insufficient resources from which most judiciaries suffer, there is a danger that support systems based on artificial intelligence are inappropriately used by judges to 'delegate' decisions to technological systems that were not developed for that purpose and are perceived as being more 'objective' even when this is not the case. Great care should, therefore, be taken to assess what such systems can deliver and under what conditions that may be used in order not to jeopardise the right to a fair trial.[22]

6. Conclusions

Before any properly functioning 'AI judge' is created, consequences revealing the full picture of potential advantages and risks of such evolution in civil proceedings needs to be reliably examined. Both full automation of legal proceedings via artificial intelligence systems taking over all functions performed by the judge and the use of AI tools as the judge's support system must remain in line with the democratic rule of law and follow provisions shaping the content and form of judicial procedure. Without the detailed analysis of the compliance of AI implementations with national, European, and international legal orders, it is completely useless to thoughtlessly implement new technological solutions or raise hasty hypotheses about the inevitability of replacing the lawyers with artificial intelligence.

My future research will include the assessment of whether in judicial proceedings conducted with the support of AI all leading principles characterizing the content and form of court procedures are respected. It will allow the evaluation of the possibility of implementing AI tools into judicial procedures:

1) without the necessity to amend the provisions of law,

2) by partial or substantial changes in the legislation (including constitutional regulations), and

3) by creating brand new fully automated (but non-judicial) solutions for settling legal disputes.

The research undertaken is aimed at complementing the efforts of AI and law researchers (focused on modelling or building artificial intelligence systems into

22 The Council of Europe Study DGI(2017)12 'Algorithms and Human Rights – Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular algorithms) and Possible Regulatory Implications', prepared in March 2018 by the Committee of Experts on Internet Intermediaries (MSI-NET), March 2018 (https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5, accessed: 20.08.2019), 8, 12.

the work of lawyers) by carrying out reliable legal and interdisciplinary analyses of the admissibility of using AI in the judiciary.

Finally, it is worth highlighting that if the technological development characterized by the creation of a well-functioning automatic legal adjudication system will get ahead of the analysis of the compatibility of such solutions with law or the assessment of the level of social acceptance for the use of artificial intelligence injustice, the consequences may be difficult to predict.

References

COENEN, F.–BENCH-CAPON, T. 2017. A Brief History of AI and Law. *AI & Law Workshop at BCS SGAI AI'17*: https://cgi.csc.liv.ac.uk/~frans/KDD/Seminars/historyOfAIandLaw_2017-12-12.pdf (accessed: 20.08.2019).

DA SILVA, N. C. et al. 2018. Document Type Classification for Brazil's Supreme Court Using a Convolutional Neural Network. *Proceedings of the Tenth International Conference on Forensic Computer Science and Cyber Law – ICOFCS 2018, São Paulo, Brazil*: https://cic.unb.br/~teodecampos/ViP/correiaDaSilva_etal_icofcs2018.pdf (accessed: 20.08.2019).

DIJKSTRA, J. J. 2001. Legal Knowledge-Based Systems: The Blind Leading the Sheep? *International Review of Law, Computers & Technology* 2: 119–128.

GOŁACZYŃSKI, J. 2010. *Informatyzacja postępowania sądowego i administracji publicznej*. Warsaw.

GOŁACZYŃSKI, J.–MĄCZYŃSKA, E. 2017. *Ochrona praw wierzycieli w Polsce. Wymiar ekonomiczny. Koszty transakcyjne. Prawne formy zabezpieczeń. Informatyzacja sądownictwa*. Warsaw.

RUSSELL, S.–NORVIG, P. 2010. *Artificial Intelligence. A Modern Approach*. Upper Saddle River.

TODD, P.–BENBASAT, I. 1994. The Influence of Decision Aids on Choice Strategies: An Experimental Analysis of the Role of Cognitive Effort. *Organizational Behavior and Human Decision Processes* 1: 36–74.

Appendix I to the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment Adopted by the Council of Europe European Commission for the Efficiency of Justice (CEPEJ) during Its 31st Plenary Meeting, Strasbourg, 3–4 December 2018.

The Civil Procedure Code in the Act of 9th January 2009 on the Amendment to the Civil Procedure Code and Other Acts (as published in the Official Journal in 2009, number 26, item 156); http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20090260156 (accessed: 20.08.2019).

The Council of Europe Study DGI(2017)12 *Algorithms and Human Rights – Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications*, prepared in March 2018 by the Committee of Experts on Internet Intermediaries (MSI-NET), March 2018 (https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5, accessed: 20.08.2019).

Acta Univ. Sapientiae, Legal Studies, **8**, 2 (2019) 191–212

DOI: 10.47745/AUSLEG.2019.8.2.04

Artificial Intelligence as an Instrument of Discrimination in Workforce Recruitment

Alessandra Miasato
undergraduate student
Universidade Presbiteriana Mackenzie, São Paulo (Brazil)
E-mail: alessandra.miasato@gmail.com

Fabiana Reis Silva
undergraduate student
Universidade Presbiteriana Mackenzie, São Paulo (Brazil)
E-mail: fabianareis.nm@gmail.com

Abstract. The purpose of this article is to reflect on the use of artificial intelligence in the process of hiring and on how biased algorithms can pose a great risk of discrimination to particular groups if artificial intelligence is not used properly with an emphasis on labour relations. Based on current research, we present the wide range of uses how AI technology can be deployed in the search for employees who satisfy the needs of employers on the labour market. The various manifestations of bias in AI implementations utilized in the field of human resources as well as their causes are presented. We conclude that in order to avoid discrimination due to either wilful programmer behaviour or implicit in the data used to train AI agents, the observance of legal and ethical norms, as outlined in tentative projects underway worldwide, is necessary.

Keywords: biased algorithms, discrimination in labour relations, artificial intelligence

1. Introduction

Discriminatory behaviours are part of the society for various reasons, many with historical origins. And when it comes to labour relations the risk of discrimination is very high due to the applicant's skin colour, sexual orientation, gender, or physical aspects in general, among others. This discrimination can occur both during the selection process for a certain job and the execution of the employment contract.

In entirely personal interviews, there is a greater risk of a candidate suffering from the prejudice any potential employer may have against them. However,

with the use of new technologies, such as artificial intelligence, the idea has been proposed that the selection process may become fairer and more objective, any analysis being limited only to the necessary features that a candidate has to conform to in order to fill the vacancy offered.

Technology has always been part of labour relations. From the 1st Industrial Revolution to the 4th and current one, technological innovations have completely changed the ways of working. Today, artificial intelligence – AI programmed by algorithms that enable the various circumstances to be analysed in seconds, and a huge amount of them to be considered at the same time – enables a greater degree of competition.

Decision-making algorithms are defined by the data the AI is initially provided, and if the content that feeds the AI responsible for the selection of a candidate is discriminatory, the result will also be discriminatory.

So, what are the impacts of a scenario when artificial intelligence is programmed with a biased algorithm to select which candidate to be hired and which not? Surely, this is a problem that the new technology has brought along and that must be analysed in order to understand how this happens and how to handle these situations.

Today, software tools exist that are able to identify the probability of a person suffering from depression or of a woman getting pregnant. Further examples are algorithms which, by using a photo, can identify if a person is gay or straight, algorithms which can tell if a person is black or white by the analysis of his or her name. These algorithms are biased and offer a very high risk of discrimination.

2. Artificial Intelligence, Algorithms, and Discrimination

The whole of society is undergoing a process of transformation in an accelerated manner, which has never been seen before. It is in the course of transforming into the information society, and knowledge provided by different technologies is increasing in ways vastly different than have been known before. This change is what is now being called the Fourth Industrial Revolution.[1]

An industrial revolution is characterized by abrupt and radical change and is associated with the emergence of new technologies that change the entirety of society, especially the political, economic, and social sectors.[2]

The First Industrial Revolution, which took place during the XVII–XVIII centuries, was the process of the mechanization of production by the use of water and steam power as sources of energy. In the late nineteenth century, the Second Industrial Revolution took place with large-scale, quick, and inexpensive

1 Schwab 2016. 160.
2 Novais 2018.

industrial output, having electricity as the main source of power.[3] However, in the 1960s, information and communication technologies marked the Third Industrial Revolution. This gave rise to the digital revolution, and with it came the computers and their continued use by society, such as the Internet and digital platforms.[4]

With the Fourth Industrial Revolution came a new era marked by an entire set of disruptive technologies such as robotics, augmented reality, big data, nanotechnology, the Internet of Things, artificial intelligence, and many others. In today's society, we have the convergence of digital, physical, and biological technologies that make this revolution and the advent of the digital era possible.[5]

2.1. Artificial Intelligence

Today, digital revolution is difficult to differentiate from the rise of artificial intelligence (AI), which is set to become part of all aspects of life. When thinking about artificial intelligence, it is almost impossible to prevent the first thing that comes to mind being a picture of a robot, like the ones any science fiction movie portrays. AI was functionally first described by Isaac Asimov, the Russian creator of the classic *I, Robot*, known as the father of robotics.[6] But AI can manifest itself in anything from weapons and autonomous cars to search algorithms.

Artificial intelligence is a growing technology in various aspects of life, and there is a certain definition of what it is. It can be said that: '[…] it is an umbrella term that includes a variety of computational techniques and associated processes dedicated to improving the ability of machines to do things requiring intelligence, such as pattern recognition, computer vision, and language processing'.[7] In other words, it is the science of mimicking some aspects of human intelligence by use of a machine.[8]

Among the many changes that the Fourth Industrial Revolution brought, there is no denying that the AI is changing the world the most. It is hard to think of something that does not involve the use of this technology or is not a result of it. Almost everyone carries in his pocket a mobile phone that uses or implements some form of AI; there are already intelligent and autonomous vehicles which drive themselves or smart homes that can perceive the lack of food in a refrigerator and make direct requests to supermarket websites, thereby shopping for/by themselves.[9]

3 Schwab 2016. 160.
4 Novais 2018.
5 Novais 2018.
6 Seiler 2019, Isaac Asimov Home Page.
7 Raso 2018. 63.
8 Borgesius 2018. 51.
9 Morgan 2014.

A disruptive technology, which causes changes in the way of living, revolutionizes the way of thinking or acting as it becomes necessary to daily life. In short, it can be said that this change comes from a phenomenon of radical transformation in the way data and information are being processed in various sectors and activities which were previously only able to utilize human labour.[10]

In this new society of information and knowledge, the word *data* is magic, assuming a central role in such a way that the control and management of data in the midst of this society confers great power upon any entity. Artificial intelligence is powered by data, information that enables it to run the task it has been given.[11] With such data, learning and discoveries can be automated by the frequent accomplishment of voluminous tasks in a computerized way; the use of this technology brings greater security.[12] AI is able to provide data analysis much faster and deeper than a human, reaching incredible precision, making it a very reliable tool.

AI provides intelligent tools, and the process of knowledge creation is improved by its use. However, for the implementation of this technology, human interference is still essential to configure the systems and give commands for performing tasks. All the commands that the AI receives, all the data that feeds the machine are made by a language called algorithms.

2.2 Algorithms and Discrimination

There are many benefits that the use of AI may provide, but, like any other technology, it is necessary to be aware of the negative aspects of its use. In this sense, many negative points can be highlighted by the use of AI, and some of them are directly linked to the language that defines the action of the machine, i.e. the algorithms themselves.

2.2.1. What Are Algorithms?

The concept of algorithms was formalized for the first time in 1936 by the definition of the 'Turing Machine' by Alan Turing, and it is regarded today as *a finite numerical sequence of executable actions which seeks a solution to a given problem by the use of accurate, efficient, and correct procedures*.[13] Algorithms now dominate daily life, providing communication, making it possible to search the Internet, identifying musical preferences, assisting in GPS location,[14] data encryption, and more;[15] there is no escape from algorithms.

10 Mendonça 2018.
11 Novais 2018.
12 Novais 2018.
13 Ziviani 2011.
14 GPS: Global Positioning System.
15 Gangadharan 2014.

First of all, an algorithm is a set of instructions or commands to perform a certain task. If this task corresponds to a simple query entered by a user in a search engine, it can be defined as an algorithm.[16] For the purposes of this study, the analysis will be based on algorithms that are computable, i.e. those that can be read and implemented by computers. Algorithms of this type are codes that a computer is able to 'read' and execute (run). In a simple way, algorithms are nothing more than 'recipes': a step-by-step showing of the procedures for solving a task. They use variables and a process to ensure the goal, and in this digital process, which involves software, decisions are taken automatically from the data that are fed to the program.

Algorithms are used by all digital services and programs and are part of everyday life for everyone. Algorithms have become an important subject in various fields of study in addition to computer science, such as law, economics, biology, and labour relations, among others, and for this reason Gillespie stated that the findings and results that are generated by an algorithm have a powerful legitimacy, equalling the statistical data that reinforces scientific claims.[17] Thus, it has been said that the results presented by algorithms present a particular type of legitimacy, and this happens in a way that often ends up being considered more reliable than decisions or conclusions made by humans, considered to be full of subjectivity.[18] In other words, the results of algorithms are expected to be cleaner, more objective and are therefore regularly perceived as more assertive. In this sense, algorithms would be synonymous with sophisticated and quasi-infallible decision making due to the strict procedures and objectivity in data analysis that they provide. However, the accuracy and reliability due to the objectivity that the use of algorithms theoretically implies cannot be the only deciding factor determining whether the decisions taken by the AI, fuelled by an algorithm, are good or bad. The statistical precision and the objectivity of search algorithms are certainly very important for decision making to be reliable, 'but it would be unwise to conclude that the subjective human knowledge is therefore useless or of less value in terms of understanding and knowledge'.[19]

Algorithms seek the results of their procedures in conditions of objectivity and clarity, something humans also tend towards. But still, it is necessary to understand how objective an algorithm can actually be – provided we can speak of the *total objectivity* of algorithms. Although algorithms have brought an immeasurable capacity in the analysis and processing of data with a swiftness never before seen in order to command compliance, the concern that everyone should have is whether biased data is provided to a machine thereby making the algorithm itself

16 Mattiuzzo 2019.
17 Gillspie 2016. 18–30.
18 Mattiuzzo 2019.
19 Mattiuzzo 2019.

biased. If used incorrectly, algorithms may be responsible for spreading prejudice and increasing inequality.

2.2.2. Discrimination

Article I, Item 1 of The International Convention on the Elimination of All Forms of Racial Discrimination says:

> The term 'racial discrimination' shall mean any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life.[20]

In a similar sense, the Canadian Human Rights Commission classifies discrimination as 'an action or a decision that treats a person or a group badly for reasons such as their race, age or disability'.[21] The Human Rights Code of Ontario, which is an anti-discrimination provincial law, defines discrimination as an unequal or different treatment or harassment that causes harm. Many are the concepts of discrimination, but all essentially translate to mean that discrimination is a way of treating people differently, taking into account their physical or personal characteristics; these differences are used as grounds to justify that different people should be treated unequally, that the same rights should not be granted to them in equal proportion. In the legal framework, a potentially discriminatory act of a positive value is lawful, i.e. treatment that is aimed at the improvement of conditions of a certain group that historically, economically, or socially suffered disadvantages and is in a vulnerable situation, when compared with other groups, is acceptable (such measures constitute *affirmative action*).[22] On the other hand, any discriminatory practice other than affirmative action is prohibited. They are considered illegitimate, arising from arbitrary treatment motivated by stigma or mainly the cultural belief that somehow people in the same situations should be treated unequally.[23]

There are several types of negative discrimination that can be listed such as discrimination based on race, nationality, colour, religion, age, sex, gender, criminal records, etc.; and that differential treatment can be given in various sectors of daily life such as in employment relationships – the object of this research.

20 United Nations. 1948. The Universal Declaration of Human Rights.
21 Canada – Canadian Human Rights Commission.
22 Moreira 2017.
23 Moreira 2017.

2.2.2.1 Discrimination in Labour Relations

Taking into account the different types of discrimination that a person can suffer and applying them to labour relations, this unequal treatment happens when people with different characteristics receive different and less favourable treatment for reasons that often have no link whatsoever to the merits or the requirements for their position.[24] Discrimination in employment relationships has always existed. Since the period of slavery, we have an unequal treatment based on racial prejudice. Inequality and the different forms of treatment of workers persisted even after the abolition of slavery, even if it has somewhat transformed after the First Industrial Revolution. Although overcoming various degrading working conditions and, in theory, overcoming various acts of discrimination in the field of the employment of workers has been a legislative priority for some time now, there are still countless discriminatory practices applied in hiring. The importance and the need to discuss discrimination in labour relations have been recognized.[25] No wonder that the International Labour Organization (ILO) created Convention 111 in 1958, defining the concept of discrimination in labour relations:

> 1. For the purpose of this Convention the term discrimination includes:
> (a) any distinction, exclusion or preference made on the basis of race, colour, sex, religion, political opinion, national extraction or social origin, which has the effect of nullifying or impairing equality of opportunity or treatment in employment or occupation;
> (b) such other distinction, exclusion or preference which has the effect of nullifying or impairing equality of opportunity or treatment in employment or occupation as may be determined by the Member concerned after consultation with representative employers' and workers' organisations, where such exist, and with other appropriate bodies.[26]

The discussion on the subject of discrimination in labour relationships remained a very important topic for the ILO, which in 1998 defined the elimination of all forms of discrimination in employment relationships as a fundamental principle of any decent work.[27] But still discrimination in this relationship is a reality – whether in the course of the employment contract or at a time prior to it, during hiring.

A great part of the hiring process that was previously done in person and was time-consuming gave way to the use of AI. In the Fourth Industrial Revolution, we no longer speak anymore of huge queues in front of companies or vacancy

24 International Labour Organization 2019.
25 Lima 2011. 18.
26 International Labour Organization 1960.
27 International Labour Organization 2016.

notices in newspapers; now we speak of digital recruitment through online platforms and selection made with AI tools. The machine being used as a selector of candidates has brought many benefits to the hiring system, but it also presents a very significant risk.

2.2.3. Biased Algorithms

As shown above, algorithms are a type of language translated into numbers, which allow a computer system to read the commands given by a programmer in order to accomplish a certain task or provide answers to a problem in an objective, clear, and timely manner. But AI can also be used to solve certain subjective issues such as deciding who should be hired for a particular company, which contract should be signed, the likelihood of recidivism of a criminal, etc.[28] On the issue of autonomous cars – as an extreme example – which are programmed with algorithms, it is the algorithm that will decide if a person will be hit or not in an imminent accident situation.[29]

The algorithms must be programmed, and to the extent that this programming is done by humans there is interference at the moment of transmitting the world's impressions to the programming.[30] And it is at that point where the algorithm suffers the interference of the programmer's moral beliefs and then rise to the term 'biased algorithms' or 'discriminatory algorithms'. When an algorithm is programmed to analyse the frequency of 'likes' from a person on the Internet and what her preferences, her musical tastes, political views are, the social events she attends, her network and more, it is possible that the data collected can end up bumping into 'sensitive' information.[31] The impacts of the use of AI can be manifold, especially when it comes to software programmed to do data analysis; and, generally, the risk of any negative results is added to the machine even before it starts to operate or even before the system being developed.[32]

There are two aspects that influence decision making made by an AI: *the quality of the data* that feeds it and *the design of the system* being used. If the data used to train an AI system is biased, the consequence of this is that the system will eventually reflect and often leverage these trends. Still, system designs using AI are created by humans who may, for example, prioritize certain characteristics or certain variables, depending on how they want the machine to behave.[33] The phrase 'garbage in, garbage out' translates this problem quite well. The term is widely used in the field of programming and means that if the data that is inserted

28 Jota 2019.
29 Jota 2019.
30 Mattiuzzo 2019.
31 Mattiuzzo 2019.
32 Raso 2018.
33 Raso 2018.

into the machine is poor, the results are equally poor.[34] In other words, feeding an AI with biased information generates biased results because the problem is at the origin of the data being used. Any trends and biases that end up being incorporated into systems operated by AI through their algorithms can be the gateway to various forms of discrimination.

As Marcelo Chiavassa[35] puts it, AI is not that different from a five-year-old child: a child is not born racist, sexist, or homophobic, but if they grow up hearing racist, sexist, and homophobic comments, there is a high probability of them reflecting the prejudices which they grew up with. The AI is not that different from a child in the sense that if this AI does not have any prejudice by itself but is still powered by algorithms that reflect the opinion of a racist, homophobic, or sexist, the consequence will be a machine reproducing these prejudices and discriminations based on the information that fed it. An algorithm does not have biases by itself, this is a characteristically human trait; so, if the software operates to discriminate against a certain group, it discriminates based on the data input received. There are numerous areas in which AI can be used, among them criminal justice, the financial sector (ranking systems), healthcare (diagnostics), the education sector, and human resources (recruitment and hiring).[36]

3. The Recruitment System

Before the advent of artificial intelligence, methods of recruitment for job openings were more personal and therefore more time-consuming because there was no other way for this process to unfold. According to Idalberto Chiavenato, recruitment is a process of locating, identifying, and attracting candidates for the organization.[37] Recruitment can be both internal and external. Internal recruitment implies filling vacancies within the company either by promotion or by intra-company transfer. External recruitment is through the search for candidates in the human resources market.[38]

Such forms of recruitment require various techniques which may involve the presentation of candidates by company employees, posters in the lobby of the company, candidate pools, visiting schools and universities, advertisements in newspapers or magazines, agencies or recruitment firms, or virtual recruitment, among others.[39] This kind of personal selection could take weeks, even months until the vacancy is filled. Over time, the labour market has become increasingly

34 Neff-Mallon 2019
35 Chiavassa 2019, Podcast – Distopia.
36 Raso 2018.
37 Chiavenato 2010.
38 Chiavenato 2010.
39 Chiavenato 2010.

competitive, and this delay in hiring a candidate was no longer an option. Competition has made this system intolerably bureaucratic and time-consuming if done manually and in person, thereby promoting automation. Now, all the time spent on endless résumés and interview analyses would no longer be necessary. Besides being time-consuming, the process of hiring people was quite likely to fail, allowing for a selection which is totally arbitrary and full of demands that reflect the personal interests of the interviewer and not the company's interests.[40] Globalization and the emergence of AI have made the whole recruitment process change. With the Fourth Industrial Revolution at a fast pace, an interface between the real world and the digital world forms, and the recruitment system cannot avoid this change. (Most recruitment is made digitally, using software, social networking, and recruitment companies that use data analysis to select candidates who best fit the company's profile, among others).[41]

At first, virtual recruitment differed little from the current one because, despite using virtual means to receive information from candidates, such as e-mails, websites, or social networks, the screening process was conducted by people. Today, AI is able to perform the selection independently, without a person analysing each résumé in turn. One of the concepts behind the development of AI in the selection of candidates, besides speed, is bringing a higher standard to these selection processes, without the ideas and beliefs of interviewers affecting the choice of candidates. With this technological advance, the selections have become impersonal and based only on data shared with companies and on the existing data on the Internet. In their recruitment, companies utilize software algorithms that define the ideal candidate for the company, with the skills and characteristics needed to fill the vacancy offered. The purpose behind these procedures, besides the possibility of analysing a large volume of applicants, is also to be able to do this much faster and with lower costs because everything is done through the analysis of CVs stored in the database and of existing résumés on the online platforms.[42] The low cost, agility, and large volume of candidates who can be analysed are a great advantage provided by AI. However, there are worrying disadvantages to this technology. The first phase of the selection process is orchestrated by an AI which is entirely impersonal because the selection is made based on algorithms which have been fed into the machine. At first, it may seem that this is an advantage because the machine would make a selection aiming to meet the relevant requirements for the vacancy to be occupied.

However, these algorithms need to be created by a person, which carries with them a number of inherent beliefs and preconceptions. Could these personal convictions and ideas be transmitted to the machine? Is it possible that AI reflects the biases of its programmer?

40 Baia 2019.
41 Baia 2019.
42 Baia 2019.

3.1 The Use of Artificial Intelligence in the Selection of Candidates

3.1.1. Positive Aspects

Companies need employees, and it is known that the whole process of recruitment and selection of candidates made by HR can be hard work. For this reason, the use of AI is much appreciated at this stage. Public and private companies have made good use of AI in the process of selection for at least two reasons: first, the ability of data analysis and the analysis of candidates and, second, that there is a growing awareness that the recruitment processes are full of implicit prejudice and discrimination, and companies believe that the use of AI would reduce much of this problem, due to the objective decisions that it would take.[43]

The revolution that the use of AI caused in hiring systems is unquestionable. As mentioned earlier, one of the most favourable points is that with this technology it is possible to analyse a large number of CVs in an infinitely shorter time than was previously expended. AI brought agility and a capacity of recruitment that fulfil the requirements set out by companies as never before. The use of AI enables the identification of the profiles of people through the data published by them on social networks, such as Facebook, LinkedIn, and others, through a system of algorithms.[44] From these profiles and the processing of such data, HR companies can more rapidly and cost-effectively identify candidates that best meet the profile of the company.

Those responsible for HR believe that the use of such software would make the hiring process more objective, less partial, and would give women and minorities a better chance, something they would not have if they were interviewed by biased human managers.[45]

AI technology with the ability to decide for itself is already used in this field for hiring employees.[46] As much as the positive aspects are attractive, attention must be paid to the problems that may arise, and the fact that these technologies are already in use facilitates the discussion.

3.1.2. Artificial Intelligence as a Tool for Discrimination

AI being used as a tool for selection of candidates is already a reality. There are several technological tools that allow the employer to have a greater control and to monitor their employees from the moment of admission, permitting

43 Raso 2018.
44 Mendonça 2018.
45 Tufekci 2017.
46 Raso 2018.

constant supervision during the term of the employment contract.[47] The use of AI to analyse the information about candidates for the vacancy, an analysis which takes place before the contractual employment relationship can be harmful to the applicant depending on how the machine has been programmed.

Outsourced HR companies, or even the internal HR from a certain company, are still responsible for the recruitment, selection, and hiring of employees. It is not relevant whether recruitment is undertaken personally or through AI implementations.[48] Regardless of the manner in which this selection of candidates is made, every business has an image of the ideal employee, and it consequently ends up creating a profile of the candidates to be considered. Is it possible for the people to whom the responsibility of hiring is trusted to objectively take the desired characteristics and skills of their candidates into consideration or do their personal opinions end up 'compromising' the search results?

Analysing HR professionals' actions based on their beliefs and opinions at the time of hiring, an application called Picture Test made by Master Communications (in partnership with the Paraná state government – Brazil) can illustrate how discrimination can take place.[49] In the framework of this research, the company responsible for the test invited professionals who work in HR. They were divided into two groups, and one of them was shown pictures of white people making daily routine activities, while for the second group the same images were shown, but with black people. As a result of this test, the majority of the responses from the people who were in the second group put the white people in a position of superiority over black people. A simple image of a white person mowing the lawn leads to the deduction of this person being the owner of the house, taking care of the garden; the same image with a black person is understood as a gardener who works for the owner.[50] This is just one example of discrimination that happens in the moment prior to hiring candidates. Companies make this kind of prejudice all the time through their representatives. Now, thinking that these same people who have these preconceptions induced by just an image are the ones responsible for defining the criteria that must be taken into consideration in formulating an algorithm that will feed a machine, the impact of such biases on the selection of candidates may be noticeable. AI carries a great risk of reflecting or even expanding existing prejudices and social biases, which would infringe one of the universal principles, that of equality.[51] It turns out that the AI systems are trained to reproduce the behaviour patterns of society in decision making, such as prejudices and human beliefs.[52] The result is a machine trained to discriminate – a machine

47 Costa 2017.
48 Kenoby 2019.
49 Brazil 2017.
50 Brazil 2017.
51 Brazil 2017.
52 Raso 2018.

that reflects social patterns of prejudice and tends to perpetuate these mistakes in every decision that is submitted until new data are inserted into it, with updates on how to make decisions (i.e. updates on new social understandings). Although the ability to change their moral perspective over time is a virtue of human beings, AI does not have this possibility.[53]

Hiring employees is in itself biased, and there is a risk that this bias can be transferred to machines. Hiring free of discrimination would certainly be a dream, and even with the use of programs that seek greater objectivity it is complicated to expect such a result. Computer systems can access a variety of information about people, including the most intimate information (religion, belief, gender, sexuality, political views, etc.), which are known as sensitive data. These programs can access this information without even revealing it and with a high degree of assertiveness.[54] The algorithms can be programmed to seek information and profile job applicants for a job. The algorithms can obtain information on social networks such as data regarding political views, religion, sexual orientation, and many other aspects that are part of the intimacy of the human being.[55] And with that information obtained, a system can be programmed to discard candidates with certain physical characteristics such as skin colour; or not even to consider them for the position.

In 2018, the Cambridge Analytica company, responsible for collecting and processing data, was accused of extracting large amounts of private information from large numbers of users of Facebook, and this data was used for political purposes. Once having access to this information, companies could create political and ideological profiles that are able to influence political views.[56] Much of the data obtained in this operation was extracted, retained, and exploited, and even though this data was 'available' to the public, the idea of an ideological profile of someone is scary. The risk that this represents is high, and when it comes to work, the vulnerability of a candidate is further accentuated by the possibility of an employer having access to a person's profile that contains all of their political views, ideology, sexual orientation, or even religious affiliation. The chance of discrimination is almost inevitable.

A study from Stanford University, the United States of America, showed that AI can deduce sexual orientation based on photographs of people's faces. An algorithm developed for this purpose was the experiment done on a dating website and showed an accuracy of 91%.[57] If a company responsible for selecting candidates has an algorithm for this purpose, by analysing images of their

53 Raso 2018.
54 Tufekci 2017.
55 Mattiuzzo 2019.
56 Wong 2019.
57 Levin 2017.

candidates, this tool can be used for anti-LGBTQ purposes. Thus, it would induce a discriminatory conduct, an algorithm being used to discriminate candidates based on a characteristic that is not related to the function that will be fulfilled.

In a lecture at TED Global, programmer Zeynep Tufekci mentioned a computer system that was developed by a friend, being able to measure the probability of postpartum depression.[58] The speaker adds that 'the results are impressive. The system provides the probability of depression months before the onset of any symptoms'. It is inevitable not to think about the positive impact that this type of innovation has on medicine. This technology would cause a revolution in the prevention of such diseases. However, the speaker herself criticizes the use of this program in the context of hiring. A company would not hire an employee if they knew they had a high probability of having depression within the next two years. There are many risks and many forms of disqualification of a candidate. 'Our artificial intelligence can fail in ways that do not fit the standards of human errors, in ways that we do not expect and for which we are not prepared. It would be terrible not get a job for which you are qualified […]'.[59]

In 2012, the Target company came under the spotlight due to an ongoing problem of using AI to identify which of their clients were pregnant, from their shopping and research habits. It was a marketing move that aimed to boost sales. The company has mapped out its clients, those who were very likely to be pregnant, and with this data collection they managed to advertise based on current and future needs. The case became known for the fact that a teenager did a search due to which the Target's system identified her as pregnant, and she received an advertisement at home with offers for pregnant women. The teenager's father did not like it and sought satisfaction from her local store in Missouri, claiming the store was encouraging his teenage daughter to get pregnant. Later, the father discovered that his daughter was really pregnant.[60]

The case mentioned may not be related to the topic in question (hiring using AI), but using the same reflexive line adopted by the programmer Zeynep Tufekci in her lecture, it is possible to perceive the risk that this type of technology brings. This same program, used to assess people's behaviours through their online activities, could identify women who intend to become pregnant in the upcoming years. The companies would not hire a woman who is likely to need maternity leave within the next 1 or 2 years.[61]

AI is already here, it is part of the daily routine, making life a lot easier but also making mistakes. Companies make use of these tools, and they know the risks that these systems may be discriminatory depending on how the programs are fed. In

58 Tufekci 2017.
59 Tufekci 2017.
60 Agostini 2012.
61 Welchering 2014.

2014, Amazon began using a computer program to hire its engineers. After a while, they realized a big problem: 'their new recruiting engine did not like women'.[62] The system used data analysis through AI to rank the best candidates. But in 2015 Amazon realized that the new system did not classify candidates for employment in a neutral way in terms of gender.

That is because Amazon's computer models were trained to vet applicants by observing patterns in résumés submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry.'[63]

In our example, the Amazon system understood that male candidates were preferable over female candidates, and this made AI discriminate against women – not based on their ability but based on gender. Later, Amazon edited the program, so it has become more neutral and no longer exhibits this conduct of discrimination between men and women;[64] but there is no guarantee that certain features will not be prioritized over others.

Discrimination in the labour market due to a résumé can start even with a simple picture. It is through the picture that an employer can get information about the physical aspects that might interest them; in the case of a racist person, black people will certainly be discarded from the selection; the same would happen if the person was sexist, discarding female candidates.[65] It was very common for companies to demand photos in the curriculum in order to have an idea about the appearance of their candidates and to see whether they 'fit' in the company's standards. Today, it is difficult to find a company that requires a photo in the curriculum because now this type of requirement is no longer appropriate[66] (although there are companies that understand that the requirement of a photo is not a discriminatory conduct).

Despite it no longer being customary to require photos, 'limiting' the ability of companies to do a preliminary analysis of the appearance of their candidates, some companies have found another way to identify some characteristics of their candidates such as skin colour. A study in the United States of America in 2004 analysing résumés showed that companies can find out the colour of the skin of their candidates through the name that appears in the résumés. Names that 'sound white' receive 50% more callbacks than candidates who have names that 'sound African American'.[67] The research brought as examples some names that exemplify what they call names that 'sound white' and names that 'sound African American': Brian and Emily are names that refer to white people, while names like Jamal and Lakisha are names that refer to African descendants.[68]

62 The Guardian 2018.
63 The Guardian 2018.
64 The Guardian 2018.
65 G1 2018.
66 G1 2018.
67 Bertrand 2004. 991–1013.
68 Bertrand 2004. 991–1013.

But for this to happen we need someone to tell the AI responsible for this analysis which names it should consider as 'good' and which it should categorize as 'bad' so as the machine can understand what curriculum vitae it should discard. The system needs objective characteristics that will make it select one or another person; furthermore, the programs responsible for making this screening are fed with algorithms that identify a number of names that, according to the person who feeds the machine, are understood as 'good names' and 'bad names' in order to dismiss all those who fall into 'bad names'.[69] The problem with this is in the algorithms that feed this AI, which are created to discriminate on the basis of training that their developers have given to these tools.

However, such discrimination does not always take place in a purposeful way. Often, at the time of programming an algorithm, the programmer may end up promoting some discrimination without realizing it. If an employer wants to hire a good employee using AI, at the time of feeding this technology they need to tell it what they see as 'good' and what they see as 'bad' to characterize their candidate. For example, if that employer understands that 'a good employee is one who is never late' and tells that to the AI, without realizing, they can discriminate against people who live in remote areas based on their addresses or people who depend on public transport. The algorithm that makes the analysis of the résumés can understand, with this information, that these people are highly possible to be delayed by their circumstances.[70] The fact that a person relies on public transport is not a delay guarantee and has no direct link with the function that the person was hired to fulfil. Even if it was not the intention of the employer or company responsible for selecting candidates to discriminate, it can happen without them noticing.[71]

Defining certain objective characteristics can be sufficient to induce discriminatory behaviour. If an employer believes that a good employee is one who is available to work on different schedules, AI can understand that people who have children do not meet the job requirements, deducing that people with children require a certain routine, a fixed timetable.[72]

AI as a hiring tool can pose risks of discrimination in many forms, and this has become of major concern to many companies that realize the damage this can cause not only to the programmers who develop the algorithm but especially to the candidates for job openings that are in an even more vulnerable situation. There needs to be concern for how algorithms are used and the purposes they have. The risks to discriminate are too high when the algorithms are biased.

69 Bertrand 2004. 991–1013.
70 Council of Europe 2019.
71 Council of Europe 2019.
72 Council of Europe 2019.

For this reason, the European Commission created the *Ethics Guidelines for Trustworthy AI* in 2019. One of the things it brings is a simple checklist that serves as a guide for creating an algorithm in order to leave it as objective as possible and free of discrimination.[73]

Certainly, the concern with the way the algorithms are created is essential, and the person of the programmer at that time is important when the intention is to keep the algorithms without discriminatory biases.

4. Ethical Concerns of Developers

As the person responsible for creating the algorithm, the figure of the programmer has great significance, especially when there is a search for an algorithm that will work as objectively as possible. Often creating a biased algorithm cannot happen on purpose or with the intention of 'hurting' someone; but even if its creation is full of good intentions, the programmer has to be aware that their algorithm can be distorted and used for other purposes. So, there are a number of ethical issues surrounding the figure of the programmer.

A former member of the American military, Chelsea Manning, also a programmer, was jailed for seven years after being responsible for one of the biggest leaks of classified information in the history of the United States of America. Even after her arrest she remained a reference in programming. In a panel at SXSW[74] 2018, the programmer showed her concerns about the creation of a code of ethics for programmers because the power that an algorithm can wield. She said it is pure deception that some tools are created with algorithms free of bias.[75] In the words of the programmer: 'the systems are biased, yes. Be it in the way the algorithms are written, either in the way they are fed with data'.[76] Chelsea then completed her statement by saying that programmers are obliged to think of the consequences of the algorithms that they create and that the same way doctors have an ethical code programmers must also have one due to the amount of power they give to algorithms and to those who control them.

Nevertheless, this concern about the ethics of programmers is not restricted to her person. Some universities in the United States as well as the European Union began to worry about the lack of a code of ethics for programmers.[77] Universities began to worry about some ethical issues behind the use of AI and have offered courses in order to warn the next generation of experts in technology about the

73 European Commission 2019.
74 South by Southwest, a.k.a. SXSW, is a festival considered one of the world's innovation events.
75 Manning 2018.
76 Manning 2018.
77 Estadão 2018.

dark side of the innovations.[78] The idea is to show students that technology is not neutral and these biases it carries generate social impacts. Professor of Computer Science at Stanford University, Mehram Sahami, points out that students should be prepared for the challenges they encounter in the future of technology, knowing how to handle it, so it does not generate negative impacts.[79]

Since new technologies can bring great changes in society, universities in the United States rushed to make students understand the risks of technology misused or used in a distorted manner. And due to the risk of these consequences being so big, the discussion about the creation of an ethical system for programmers is presented as a solution.[80] Medical professionals have a code of ethics, and likewise these universities argue that a 'unique' code of ethics should be created for programmers.

Because of the concern that programmers may (even inadvertently) transmit their ideologies, beliefs, and morals to the algorithm that is created, the guide created by the European Commission suggests that when creating an algorithm the programmer should answer some questions in order to ensure that the AI will work in a way which avoids biases without risk of discrimination.[81] The 'guide' to avoiding discrimination and to having no biased algorithms consists of four main groups of questions that developers should answer to be sure about the objectivity of their algorithms. They are questions like: 'Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?' or 'Did you consider diversity and representativeness of users in the data? Did you test for specific populations or problematic use cases?[82]

The use of biased algorithms that have a high risk of discrimination is dangerous, and when discrimination takes place there must be accountability. Nonetheless, because of the difficulty of defining who is 'guilty' for the act and due to the fact that this is a recent issue in the world, there is still no definition of who should be blamed for this. Furthermore, the intention of the European Union to create the ethical guidelines and checklist for programmers was to minimize these problems and the negative impacts that the misuse of an AI fuelled by biased algorithms can cause. This is a small solution – until a more efficient method to 'control' these technologies does not arise –, but it can have a positive impact on the use of these technologies.

78 Estadão 2018.
79 Estadão 2018.
80 Estadão 2018.
81 European Commission 2019.
82 European Commission 2019.

5. Conclusions

There are many benefits brought by the use of AI. So many that it is hard to think of any day-to-day activity that is not controlled by an algorithm or that does not have a technology behind it. While the software used by companies has brought more agility to the recruitment process and enabled greater competition among the candidates, the use of AI in the process ended up opening a door for discrimination.

When talking about an entirely personal selection process, i.e. without any manifestation of technology, it is hard to imagine that this choice of 'perfect candidate for the job' is free of discrimination. When it comes to making that choice with AI, the discrimination of candidates still happens.

Although a machine is responsible for selecting the résumés, there is programming behind it, there are a number of guidelines that the programmer gave to the AI to be able to decide which candidate is the best. As much as the intention with the use of this technology is to select candidates based on their ability to fulfil the job, we cannot ignore the fact that those who do the programming or who decide the form in which this algorithm will be programmed is done by a person.

The risks of discrimination with the use of AI are many, and the main reason for this is the fact that behind an algorithm there is the possibility of a person with prejudices that will reflect the machine's decisions.

A homophobic employer would not hire a person who is not straight to work in their company; the same goes for sexist or racist employers. And with technological advances, the possibility of obtaining such personal information of their candidates is becoming easier. With algorithms being created in order to identify LGBTQ people from pictures, programs that can identify if a person is black or white from the name, or even software developed to predict whether a woman plans to become pregnant, the risk of some groups of people being discriminated against increases.

Taking into account the vulnerability of a worker in the position of a candidate for a job vacancy, it is possible to see the benefits of a bias-free algorithm that does not present risks of discrimination. But this is not the reality – since algorithms cannot be neutral, there must be a concern with the use of these technologies.

The negative impacts from the inappropriate use of an algorithm are manifold. The programmers should care about the quality of the algorithms they create, the purpose of their programs, and the intentions behind their use.

Discrimination followed by accountability at this point of the relationship is a very sensitive issue since there is still no contractual relationship, and the AI used as an instrument of candidate selection as well as the determination of accountability have become even more obscure. There is not yet any standard that assigns responsibility for this act, which is why the discussion on the impacts of using biased algorithms for this purpose becomes increasingly necessary.

Companies need to raise awareness about the risks of discrimination and how these technologies can be dangerous for certain groups. And while there is no type of agreement or rules that determine how AI should be used in order not to discriminate and how accountability for discrimination occurs, it is important that the subject be debated repeatedly and that a number of tests be made to ensure that AI will not be discriminatory in any way. More than ever, there is a need for debate on the subject and for research to be done to consider both legal and technical issues surrounding technology. If the discussions and tests are made as soon as possible, AI can be best enjoyed, while the risks of unfair discrimination are minimized.

References

AGOSTINI, R. 2012. A nova indústria da espionagem explora o consumo. *Exame*: https://exame.abril.com.br/revista-exame/a-nova-industria-da-espionagem/.

ALGORITMOS e inteligência artificial – repercussões da sua utilização sobre a responsabilidade civil e punitiva das empresas. Newspaper JOTA. 2019. https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/algoritmos-e-inteligencia-artificial-15052018#sdfootnote1sym.

AMAZON Ditched AI Recruiting Tool That Favored Men for Technical Jobs. *The Guardian*. 2018.

BAIA, C. 2019. Como foi a revolução da Inteligência Artificial no Recrutamento e Seleção. *GUPY*: https://www.gupy.io/blog/inteligencia-artificial-no-recrutamento-e-selecao.

BERTRAND, M.–MULLAINATHAN, S. 2004. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review* 4: 991–1013.

BORGESIUS, F. Z. 2018. *Discrimination, Artificial Intelligence, and Algorithmic Decision-Making*. Council of Europe: https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 51.

CHIAVASSA, M. 2019. *Algoritmos enviesados no processo de recrutamento*. Podcast: Distopia.

CHIAVENATO, I. 2010. *Iniciação à administração de recursos humanos*. Manuele.

COSTA, A. D.–GOMES, A. V. M. 2017. Discriminação nas relações de trabalho em virtude da coleta de dados sensíveis. *Scientia Iuris* 2: 214–236.

COUNCIL OF EUROPE. 2018. *Discrimination, Artificial Intelligence, and Algorithmic Decision-Making*. https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73.

*DISCRIMINAÇÃO no mercado de trabalho pode começar por fotos em redes soc*iais. 2018 *G1*. https://g1.globo.com/economia/concursos-e-emprego/noticia/

2018/08/17/discriminacao-no-mercado-de-trabalho-pode-comecar-por-fotos-em-redes-sociais.ghtml.

EUROPEAN COMMISSION. 2019. *Ethics Guidelines for Trustworthy AI*. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477.

FERREIRA, A. B. H. 2010. *Dicionário Aurélio de língua portuguesa*. Positivo.

GANGADHARAN, S. P. 2014. *Data & Discrimination*. Open Technology Institute.

GILLSPIE, T. 2016. *Digital Keywords – A Vocabulary of Information Society and Culture*. Princeton.

INTERNATIONAL LABOUR ORGANIZATION. 1960. *Convention concerning Discrimination in Respect of Employment and Occupation – No. 111*. https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_INSTRUMENT_ID:312256.

2016. *A Vision Statement by Guy Ryder, Director-General of the ILO*. https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_007913/lang--en/index.htm.

2019. *Discrimination in the Workplace*. http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---declaration/documents/publication/wcms_decl_fs_95_en.pdf.

LEVIN, S. 2017. New AI Can Guess Whether You're Gay or Straight from a Photograph. *The Guardian*: https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph.

LIMA, F. A. 2011. *Teoria da discriminação nas relações de trabalho*. Elsevier.

MANNING, C. 2018. Presentation at SXSW Event in Austin, USA.

MATTIUZZO, M. 2019. *Algorithmic Discrimination: The Challenge of Unveiling Inequality in Brazil*. University of São Paulo.

MENDONÇA, A. P. 2018. Inteligência artificial – recursos humanos frente as novas tecnologias, posturas e atribuições. *Revista Acadêmica*: https://eumed.net/rev/ce/2018/4/inteligencia-artificial.html.

MOREIRA, A. J. 2017. *O que é discriminação?* Casa do direito.

MORGAN. J. 2014. A Simple Explanation of 'The Internet of Things'. *Forbes*: https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#4d96cffe1d09.

NEFF-MALLON, N. 2019. *Garbage in, Garbage out – Machines Learn to Discriminate from Discriminatory Data*. Extract Systems.

NOVAIS, P. 2018. Inteligência artificial e regulação de algoritmos. *Diálogos*: https://www.sectordialogues.org/publicacao/inteligencia-artificial-e-regulacao-de-algoritmos.

RASO, F. 2018. *Artificial Intelligence & Human Rights – Opportunities & Risks*. Cambridge.

RECRUITMENT and Selection Software – Kenoby. 2019. Recursos Humanos – Tudo o que você precisa saber sobre o RH.

SCHWAB, K. 2016. *The Fourth Industrial Revolution*. Currency.

SEILER, E. 2019. *Isaac Asimov home page*. http://www.asimovonline.com/asimov_home_page.html.

TESTE de Imagem. https://www.youtube.com/watch?v=xvDcD3_y5EQ.

TUFEKCI, Z. 2017. *Machine Intelligence Makes Human Morals More Important*. During a presentation at TED Global, New York, United States.

UNITED NATIONS. 1948. *The Universal Declaration of Human Rights*.

UNIVERSIDADES dos EUA tentam trazer ética dos médicos para programadores. *Estadão* 2018.

WELCHERING, P. 2014. *Google investe em inteligência artificial para prever comportamento humano*. https://www.dw.com/pt-br/google-investe-em-intelig%C3%AAncia-artificial-para-prever-comportamento-humano/a-17410656.

WHAT IS DISCRIMINATION? https://www.chrc-ccdp.gc.ca/eng/content/what-discrimination; https://www.hrlsc.on.ca/en/what-is-discrimination.

WONG, J. C. 2019. *The Cambridge Analytica Scandal Changed the World: But It Didn't Change Facebook*. The Guardian.

ZIVIANI, N. 2011. *Projeto de algoritmo – com implementações em java e C++*. Cengage Learning.

Acta Univ. Sapientiae, Legal Studies, **8**, 2 (2019) 213–230

DOI: 10.47745/AUSLEG.2019.8.2.05

Liability for Intelligent Robots from the Viewpoint of the Strict Liability Rule of the Hungarian Civil Code*

Réka Pusztahelyi
Associate Professor
University of Miskolc, Faculty of Law, Miskolc
E-mail: jogreka@uni-miskolc.hu

Abstract. The European Parliament resolution of 16 February 2017 on Civil Law Rules on Robotics proposed that the strict liability and the risk management approach are alternative legal instruments to achieve the goals set out by this instrument. The evolution of strict liability is parallel with technological change; our question here is whether the elaborated rules are appropriate for managing new problems. For establishing accountability, the question arises: who is to be held liable for damages and based on which form of liability? Setting aside the issues of product liability and setting aside the independent liability of the most sophisticated autonomous robots having 'electronic personality', this essay concentrates on liability questions of the user, and it examines the strict liability rules instituted by the Hungarian Civil Code and their application in practice. According to the results of our previous research, the judicial practice regarding the general clause of liability for dangerous activity (Section 6:535. HCC) is quite flexible and covers the liability issues of damage caused by artificial intelligence. We observed also that the criterion 'dangerous' means less and less risk of damage within normal circumstances, and this statement of fact in practice also successfully competes with other strict liability rules (i.e. product liability for malfunctioning medical devices, liability for dangerous animals, etc.). The capacity of the 'keeper' or 'operator' of the robot and the emerging new types of risks are also touched upon.

Keywords: strict liability, dangerous activity, artificial intelligence, extra-contractual liability, capacity of operator, Hungarian Civil Code

* This research was supported by project no. EFOP-3.6.2-16-2017-00007, with the title *Aspects on the Development of Intelligent, Sustainable and Inclusive Society: Social, Technological, Innovation Networks in Employment and Digital Economy*. The project was supported by the European Union, co-financed by the European Social Fund and the budget of Hungary.

1. Introduction

The European Parliament (EP) resolution of 16 February 2017 on Civil Law Rules on Robotics lays down as principle that the future legal solution should not limit the forms of compensation which may be offered to the aggrieved party on the sole grounds that damage is caused by a non-human agent (i.e. robot). The title of this essay uses the notion 'robot' in the meaning of artificial intelligence, which is embedded in hardware devices (e.g. advanced robots, autonomous cars, drones, or Internet of Things applications).[1]

The possible application and appearance of AI systems is manifold.[2] That means it is hard to elaborate one general clause for liability for damages which would be appropriate for all cases and for all AIs. For example: special issues are emerging from the appearance of autonomous cars on the roads because these devices are sophisticated combinations of sensors and AI software, where the latter is dedicated exclusively for this special purpose. In addition, national road traffic liability rules are like a quilted EU carpet, where only the MID (Motor Insurance Directive) provides some uniformity not on the level of liability for damages but on the level of liability insurance. Autonomous cars being a kind of motor vehicle, the danger posed by an autonomous car is mainly attributable to the operation of an engine. However, under certain circumstances, other elements of risk would also appear as new threats emerge from AI technology.

The above-mentioned EP resolution proposed that the strict liability and the risk management approach should be alternative legal instruments to achieve the above-mentioned goal. In our opinion, risk management systems, regardless of their typology (self-insurance, compulsory liability insurance, or non-fault systems), should not be able to adversely affect the rules of civil law liability, as pointed out in the relevant literature.[3]

In order to draw up some solutions for the question 'Should new rules be introduced?', we should briefly touch upon similar legal problems concerning steam engines and automobiles which appeared on the roads more than a century ago[4] and which have shaken the classic liability systems based on the subjective criterion of fault. The evolution of strict liability is parallel with technological

1 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 on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM(2018) 237 final.

2 According to the EU Commissions Communication, Artificial Intelligence for Europe (SWD(2018) 137 final), AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems), or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones, or Internet of Things applications).

3 Fiore 2017, Wagner 2006. 277–299.

4 Kolosváry 1908.

changes; our question is here whether the elaborated rules are appropriate to manage new problems.[5]

It is obvious that strict liability rules may hinder the process by which new intelligent devices become part of everyday life either as new products on the market (economic benefits) or as useful aid for elderly people with disabilities (social benefits).[6] But the uncertainty of the regulation is also to be avoided. It should also be mentioned that the fragmentation and lack of harmonization of national liability regulations could also slow down the development of EU robot liability law.[7]

For establishing accountability, the questions to be answered are: on what do we base the liability and who is to be held liable for damages. Setting aside the issues of product liability and setting aside the independent liability of the most sophisticated autonomous robots having an 'electronic personality', this essay concentrates on liability questions of the user, and it examines the strict liability rules provided for by the Hungarian Civil Code (HCC) and the adjacent judicial practice. According to the results of our previous research,[8] the judicial practice regarding the general clause of liability for dangerous activities[9] is quite flexible and able to cover the liability issues of damage caused by artificial intelligence. We also observed that the criterion 'dangerous' means less and less risk of damage within normal circumstances, and this statement of fact in practice also successfully competes with other strict liability rules (i.e. product liability for malfunctioning medical devices, liability for dangerous animals, etc.). This essay will also touch upon the identity of the 'keeper' of the robot (registered or not, owner or not), the casual link between operating a robot, and the damage caused.

2. Technological Development and Strict Liability

Technological development brings changes in society gradually. Nowadays, devices with weak or narrow AI are being utilized, and, according to conservative estimates, the strong or general AI is expected to emerge on the market around 2040. General AI is a software application that exhibits analytical, decision-making, and learning abilities similar to those of humans.[10] At the first stage, where we are currently, traditional liability rules, especially strict liability, serve as a bridge between the concepts of traditional civil liability and other, innovative concepts. In our opinion, the gradual development of artificial

5 Martin-Casals 2014.
6 Richards–Smart 2016.
7 Wagner 2018.
8 Pusztahelyi 2018b.
9 Section 6:535 HCC.
10 Artificial Intelligence for Europe SWD(2018) 137 final.

intelligence entails innovative thinking and gradual development of liability law. We agree with what the relevant literature pointed out, that the adaptation of a strict liability rule intended to cover all kinds of uses of artificial intelligence – without any regard to the characteristics of the applications or to the specific nature of the sector – is not preferable and would be excessive.[11] Furthermore, taking graduality into consideration, it seems premature to establish liability rules in this present level of development. However, certain fields of AI utilization have already reached a high degree of technological progress, where the application of traditional liability rules raises a number of acute problems even now.

One of these fields is constituted by the issues emerging from the roll-out of autonomous cars. An autonomous car is a kind of motor vehicle.[12] According to Hungarian tort law, its operation is deemed as a dangerous activity which triggers strict liability. Generally speaking, in Europe, damage caused by motor vehicles is covered by compulsory liability insurance.[13] In continental European legal systems, risk-based liability prevails[14] although some parallels exist with the fault-based liability rule. Strict liability is replaced or at least supplemented by other compensation systems, too. It means that there are well-established compensation systems in European Member States, within the common framework of the Motor Insurance Directive.[15] The broad judicial interpretation[16] of the scope of the Directive amplifies the strictness of the liability rules as liability insurance law and tort law interact on the level of practice.[17]

11 Borghetti 2019. 72.

12 The practice recommendation issued by the Society of Automotive Engineers (J3016-2018), *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, is accepted worldwide among automotive manufacturers.

13 Directive 2009/103/EC of the European Parliament and of the Council of 16 September 2009 relating to insurance against civil liability in respect of the use of motor vehicles, and the enforcement of the obligation to insure against such liability (MID).

14 Karner 2018.

15 The amendment of the MID is adopted. One of its goals is to adjust the motor insurance system to new technological developments (autonomous or semi-automated cars, other electric vehicles, e.g. Segway, e-bikes). See COM(2018) 336 final (Brussels, 24.5.2018) 2018/0168 (COD) *Proposal for a Directive of the European Parliament and of the Council*.

16 A number of judgments of the Court of Justice of the European Union [in the Vnuk case (C-162/13), the Rodrigues de Andrade case (C-514/16), the Torreiro case (C334/16)] have clarified the scope of the Directive. According to the Vnuk judgement of September 2014, the third-party motor liability insurance obligation in Article 3 of the Directive covers any activities consistent with the 'normal function' of a vehicle, regardless of the location where the vehicle is used. The Rodrigues de Andrade judgement of 28 November 2017 clarified that only the 'normal use of the vehicle as a means of transport' and 'irrespective of the terrain' should be covered by third-party motor liability insurance, excluding accidents where the vehicle was used for exclusively agricultural use. See COM(2018) 336 final (Brussels, 24.5.2018) 2018/0168 (COD) *Proposal for a Directive of the European Parliament and of the Council*.

17 Baker 2005. 3–4.

The study commissioned by the European Parliamentary Research Service with the title *A Common EU Approach to Liability Rules and Insurance for Connected and Autonomous Vehicles*[18] determines four policy options to address the current shortcomings of the EU liability framework. These are the *status quo* (Option 1), *reform of the Product Liability Directive* (Option 2), *reform of the Motor Insurance Directive* (Option 3), and the *introduction of new EU legislation* as well as setting up a no-fault insurance framework for damages resulting from AVs (Option 4). The unification of the strict liability rules of Member States applied to artificial intelligence can be imagined within Option 4, however, with the assistance of a no-fault insurance framework. At the other end of the scale, Option 1 leaves the legislation untouched and places the onus of determining rules for liability on judicial practice, on a case-by-case basis.

In our opinion, in the first phase of the development of AI systems, the latter option would be a proper choice until the moment when the proliferation of artificial intelligence will facilitate the collection of enough comparable and reliable data to decide whether these new artificial agents cause much less damage than a human driver or not and to draft the new EU legislation which can provide appropriate compensations for victims.[19] These two options are applicable for all types of artificial intelligence, but the applicability of Option 2 and Option 3 is restricted to special utilization sectors. Especially Option 3, the reviewing of the MID, is set out with connected, automated, and autonomous vehicles in mind. Option 2, the reform of the PLD, would be a good solution, but its application has its limits. The rules of product liability, such as the concept of *product*, the *burden of proof* on the injured person, or the *causal link* between defect and damage, may not be appropriate even after a reform. For example, the defectiveness of the algorithm is hardly detectable at the time of the market release, or it may occur later during a software update or as a result of machine learning or even cybercrime.[20]

To sum up, the Hungarian legislator has two options. The first one is to maintain the status quo. In this scenario, the liability rule for dangerous activities will be applied. Issues emerging from this solution will be discussed in the following. In the case of autonomous and automated vehicles, this strict liability serves the functioning of the compulsory liability system of motor vehicles. The other possible way is a more drastic option, and it depends on the strategy the EU will eventually opt for. In case there will be no special rule, or not for all economic

18 A common EU approach to liability rules and insurance for connected and autonomous vehicles. European Added Value Assessment Accompanying the European Parliament's legislative own-initiative report (Rapporteur: Mady Delvaux) PE 615.635.

19 Borghetti 2019. 66.

20 Borghetti 2019. 72.

sectors where AI is applied, the Hungarian legislator would create new rules for the utilization of AI.

In our opinion, liability issues will be left in Member State competence for a long time, with the exception of product liability, sustaining the traditional national tort law regimes. Only certain fields of AI applications require special sectoral regulation on the EU level.[21] It is noted that the national regulation relating to automated or autonomous cars among European countries had begun to flourish.[22] The sector of automated or autonomous driving systems requires special liability rules. The question is whether there is a necessity and a possibility at the level of the EU to broaden the scope of regulation of the substantive liability law in order to harmonize traffic law as a whole. Alternatively, the national road traffic law systems may remain untouched as well as the compulsory liability insurance schemes. We emphasized these questions above to show that the application of artificial intelligence with the special issues emerging from that is only one factor in determining the future issues of liability. Our other goal is to illustrate that some special characteristics, even within the operation of an autonomous car, will put the traditional liability rules as well as the risk-based rules to the test.

Setting aside these questions, in the following, the Hungarian strict liability rule currently in force for dangerous activities will be presented. What advantages could unfold that provide a remarkable potential for this statute now, and can it be sustained for the future?

3. Rationales of the Risk-Based Strict Liability

The Commission Staff Working Document called *Liability for Emerging Digital Technologies*[23] Accompanying the Communication *Artificial Intelligence for Europe* (COM (2018) 237) differentiates between *fault-based* and *risk-based* extra-contractual liability regimes. This Document emphasizes that the risk-based strict liability rule is widespread in Europe, but in manifold forms. It highlights the speciality of the Hungarian rule for strict liability for dangerous activity. As it is a 'Generalklausel' (blanket clause),[24] it does not specify which activities trigger the strict liability. 'It is often said that fault liability is attributable to corrective/

21 See also. 'It should be discussed whether that intervention should be developed in a horizontal or sectorial way and whether new legislation should be enacted at EU level.' Commission Staff Working Document *Liability for Emerging Digital Technologies* Accompanying the Communication *Artificial Intelligence for Europe* (SWD(2018) 137 final), 21.

22 Juhász 2018.

23 Commission Staff Working Document *Liability for Emerging Digital Technologies* Accompanying the Communication *Artificial Intelligence for Europe* (SWD(2018) 137 final).

24 According to Tamás Lábady, the liability rule for dangerous activity is the most general specific case of liability. Lábady 2014. 2268.

commutative justice (justitia commutativa), and risk-based liability, by contrast, is attributable to distributive justice (justitia distributiva).'[25] In other words, the strict risk-based liability rule strengthens victims' positions to claim damages, and it serves the compensatory aims of tort law.

The historical approach to liability for hazardous activities as resulting from the literature of that time shows that the concept of causality was widely accepted as the rationale for strict liability. According to Gyula Dezső, the rationale of causality can serve as a common theoretical ground for all strict liability rules (liability without fault).[26] According to Béni Grosshmid, the *quasi delictum* is a presumptive *delictum*, in the case of which the legislator disregards the duty to prove fault (and does not even allow its proof).[27] According to Gyula Eörsi, the obligation for compensation of the damage caused by hazardous operations encompasses three fields: the field of damage caused by subjective fault, the field of damage caused by conduct without subjective fault, and the mere indemnification without even holding someone liable for it.[28] Géza Marton stressed risk allocation at a societal level as a function of liability for damages. He emphasized that the rationale of the *Gefährdungsprinzip* (the endangerment principle), *risque creé* (i.e. the created risk in and of itself) is not the best way to establish strict liability. He instead placed the principle of equity in the focus.[29] If the rationales which are listed by Géza Marton as theorems of strict liability – besides the principle of prevention (deterrence) – are surveyed, the principle of *aktive interesse (within the same conceptual scope as the principle of cuius commodum eius periculum)*[30] and the principle of *societal distribution of damage* should also be mentioned. The principle of societal distribution of damage means that the person who must bear the burden of the risk of damage is the one who can better distribute the loss suffered among members of society than the victim could.[31]

Another rationale is based on the concept of permission: someone who is permitted to use a particularly dangerous thing for her own advantage should equally bear the associated risks.[32]

25 Karner 2018. 368.
26 Dezső 1932. 192.
27 Grosschmid 1900. 398.
28 Eörsi 1972. 67.
29 Marton 1931.
30 The law may attribute liability to the person that carries out the activity because this person has created a risk, which materializes in some damage and at the same time also derives an economic benefit from this activity. See Working Document, 8. (Briefly: the one who takes the advantage also shall bear the risk.)
31 '…every enterprise has to bear its own costs, damages included, or it has no place under the sun.' See: Eörsi 1975. 215–235. For a law and economics approach, see Wagner 2018.
32 Karner 2018. 368; see also: Ehrenzweig 1966. 1454–1455.

4. Liability for Dangerous Activity in the Hungarian Civil Code[33]

Under the provision of HCC Sec. 6:535. subs 1: 'A person who pursues an activity that is considered dangerous shall be liable for any damage caused thereby.' The Hungarian Civil Code applies a *blanket clause* (Generalklausel) to establish strict liability for dangerous activities, leaving undefined which activities are assumed to be dangerous. Thus, the category is left undefined by legislation (except for certain special activities, e.g. for nuclear power generation and referring rules, e.g. keeping of dangerous animals or pollution of the environment) but is determined instead by jurisprudence on a case-by-case basis (e.g. use of motor-propelled vehicles or machinery, of explosive or toxic materials, of firearms, etc.).

In European countries, this general approach to dangerous activities is quite unique, only a few countries (e.g. the Italian and Portuguese Civil Code) have risk-based general strict liability rules covering a large scale of dangerous activities. Although the relevance and appropriateness of the classical differentiation between subjective fault-based and objective (no-fault) liability are fading, in this viewpoint, the strict liability rule is nearly as abstract as the general fault-based rule. Every new risk of technological development can be subjected to this rule as the judicial practice interprets the given case often with the help of analogy.[34]

In European countries where special strict liability rules to specified types of dangerous activities prevail, tort law reforms have been drafted – as a tendency in the last decades[35] – with a general rule for dangerous activities, because of the gaps in the fragmented regulation. Even the application by analogy of special rules is accepted by certain national legal systems.[36]

In our opinion, the rule of liability for dangerous activity serves as a general strict rule which can compete with other strict liability rules such as product liability (HCC 6:550–6:559) or liability for building damages (HCC 6:560–561). An examination of the Hungarian judicial practice shows that this characteristic prevails even against the applicability of the product liability rule[37] in spite of the fact that the Directive (PLD) imposed on Member States an obligation for exhaustive harmonization.[38]

33 Act No. V of 2013 (a Polgári Törvénykönyvről).

34 Pusztahelyi 2018a.

35 For example: Griss–Kathrein–Koziol 2006, Reischauer–Spielbüchler–Welser 2006, Huguenin–Hilty 2013.

36 Battesini 2005. 7–9, van Dam 2013.

37 In a case resolved by the Supreme Court of Hungary, No. BH2005.251, and in another case, No. BDT2016.3459, the medical devices malfunctioned and caused serious injuries to the patients.

38 'According to the CJEU, if the claim falls under the scope of product liability, the national court is prevented from applying parallel regimes of national law, even if the alternative could be more beneficial for the victim.' See Menyhárd 2017. 13–18. We agree with Attila Menyhárd, who

Studying the interactions between liability for hazardous operations and liability insurance, authors emphasize the flexibility of this general clause, whose interpretation allows the statute to be adapted to the concepts and institutions of insurance law. For example, in the field of motor vehicle insurance, the operation of a motor vehicle also covers the case when the engine is not running but the driver causes damage to other persons (by opening the door). In our opinion, from a theoretical viewpoint, this flexibility is not the best way to provide the proper coordination of the two systems, but this feature serves the applicability of the liability rule in challenging and modified circumstances created by new liability insurance schemes for AI technology. So, the judicial practice regarding the general clause of liability for dangerous activities is quite flexible[39] and able to cover the liability issues of damage caused by artificial intelligence.

Nevertheless, it should be highlighted here that domestic extra-contractual liability rules (not only fault-based but even strict ones) with their complexity generally do not facilitate victims' claims to compensation because of the burden of proof which is placed on the victims themselves or due to evidential difficulties. From this viewpoint, the special liability rule for dangerous activities has a great potential compared with product liability, for instance. In addition, the Hungarian extra-contractual liability regime does not exclude that a fault-based and risk-based liability could exist alongside one another. The competing contractual and extra-contractual claims will be discussed later.

In the viewpoint of the Law and Economics approach, Gerhard Wagner states: 'It is required to keep an eye on the different components that together represent the costs that accidents impose on society. One important component is the cost that accidents impose on victims, another is the cost that potential injurers incur for taking care, i.e. for taking precautions that prevent accidents from occurring.'[40]

He emphasized, that: '[T]he administrative costs of operating a liability system must not be ignored. Liability rules should not be based on elements that are difficult and therefore costly to establish in legal proceedings before a court or in settlement negotiations with responsible parties or their insurers.'[41]

It should be also mentioned here that the administrative cost can be reduced not only by shifting the burden of proof but also by simplifying the proceedings for enforcing claims for damages (e.g. in a collective redress procedure). The claim for damages grounded on liability for dangerous activity triggers relatively low administrative costs.

remarks that the national courts are obviously reluctant to draw the consequences of the maximum harmonization established by the CJEU in cases C-154/00 Commission of the European Communities v Hellenic Republic and C-183/00 María Victoria González Sánchez v Medicina Asturiana SA. See also Whittaker 2014. 175–176.

39 Pusztahelyi 2018b. 3–8.

40 Wagner 2019. 31.

41 Wagner 2019. 31.

But how can we determine the dangerousness of the activity? It is still a challenge to determine the degree of the danger posed (how high or extraordinary it is). The Hungarian blanket clause provides the opportunity for the judge to determine and assess this in a way which is appropriate to the case. In judicial practice, in order to assess the dangerous nature of an activity, one should consider the characteristic features of the device applied in the course of the activity and the potential consequences of the events triggered by this activity. The issue should be assessed on a case-by-case basis. Whether a slight abnormality occurring under normal conditions of use can cause damage in a disproportionately wide range or disproportionately large amount should also be held in view.[42]

However, the level of potential hazard at which the court finds the activity as dangerous is decreasing. It means that the core element of this statute would be assessed in most cases. This characteristic also helps the applicability of this liability rule.

Both causality and the conditions for exoneration from under responsibility display special features. As far as the causal link is concerned, the prevailing legal opinion emphasizes some presumption of the causality. *If the material harm (damage) is one of the normal or predictable consequences of this activity, it falls under the scope of inherent danger of this hazardous operation. Therefore, the causal link is presumed to exist*, with the exception of situations when there are several different possible causes which could have contributed to the damages occurring. This causal link extends also to the external causes which enter the scope of hazardous operation when the keeper/operator is obliged to prevent the negative impacts. In our opinion, this causality is also reflected by the exoneration rule: 'Where such person is able to prove that the damage occurred due to an unavoidable cause that falls beyond the realm of dangerous activities, he shall be relieved from liability' (HCC Sec 6:535 subs 1).

Finally, we should mention a fundamental change introduced by the new Hungarian Civil Code. The provisions of HCC draw a line between the two regimes of contractual and non-contractual liability for damages, and exclude parallel compensation claims: 'The obligee shall enforce his claim for compensation against the obligor in accordance with the provisions of contractual liability even if the obligor's non-contractual liability also exists.' – i.e. the principle of non-cumul.[43] It is essential to establish whether one of the contracting parties can cause damage to the other party irrespective of their contractual relationship. Furthermore, it is important to identify the legal grounds of claims for damages that are not related to non-performance or performance of a contract.

42 BDT 2010. 2358.

43 Section 6:145 HCC: The obligee shall enforce his claim for compensation against the obligor in accordance with the provisions on liability for damages for loss caused by non-performance of an obligation even if the obligor's non-contractual liability also exists.

Judicial practice in the field of the rule of non-cumul is still taking shape. Nonetheless, it can be stated that the rule is likely to be interpreted very strictly, that is, the contractual relationship will, for all intents and purposes, exclude the victim's claims on the ground of non-contractual liability.[44]

5. Operation of AI Systems as Dangerous Activity

5.1. Special Sources of Danger in AI Systems

Fear of an unknown, non-assessable or unavoidable risk emerging from activities carried out by others is the typical key element for establishing strict liability. The main question is what type of risk could emerge from AI technology, especially when the risk is rather immaterial and could result in pure economic losses (for example, in the case of using automated message systems for contract formation). In these situations, the strict rule of liability for dangerous activity shows its imperfection as the judicial practice defines danger as a possibility of suffering significant material harm. In our opinion, if this strict liability rule is desired to be applicable on a wide scale for damages caused by AI technology, the first requirement is to extend its conceptual scope to the significant risks of immaterial harms, too.[45]

In order to define and to assess the risk triggered by the operation of an AI system, one should pay attention to the above-mentioned study of the European Parliamentary Research Service: 'A common EU approach to liability rules and insurance for connected and autonomous vehicles'.[46] The legal problems emerging from autonomous cars shall constitute a testbed for lawyers in the field of application of AI technology in the retail sector on which to study liability and insurance problems. As an autonomous system, the AV shall be able to make hundreds of decisions per minute in order to cope with dynamic traffic conditions.

According to the 2018 study compiled by the EPRS, *four main categories of risk* relating to the liability issues associated with AVs are likely to emerge or become significantly more prominent with the mass roll-out and use of the AV. These new risks include:

(1) risks relating to the failure of the operating software that enables the AVs to function,

44 Judit Fazekas showed that the principle of non-cumul could endanger the enforcing of the right to damages based upon the rule of product liability. Fazekas 2017. 29. See also: Pusztahelyi 2016. 60–78, Fuglinszky 2017. 114, De Graaf 2017. 701–726.

45 For example, in the case of electronic banking, smart contracts, etc.

46 http://www.europarl.europa.eu/RegData/etudes/STUD/2018/615635/EPRS_STU(2018)615635_EN.pdf.

(2) risks relating to network failures,
(3) risks relating to hacking and cybercrime, and
(4) risks/externalities relating to programming choice.[47]

These new risks are added to the classical ones such as human error or the malfunction of the device. We can add one more source of risk to this list. This risk comes from the situations of mixed traffic where driverless and traditional human-driven cars participate in the same traffic, affecting the behaviour of each other. So, the interaction and any forms of good or bad communication between human and AI participants generates a new risk never experienced before,[48] which will be well in excess of the inherent hazards of a traditional traffic situation.

As we can see, software failure, or the 'bad' choice of the AI will occur within the scope of the activity. According to Hungarian judicial practice, network failures also fall *within* the scope of activity. Why should this concern us? Because the operator can exonerate themselves from under liability only if they prove that the damage occurred due to an *unavoidable* cause that falls *beyond* the realm of the dangerous activity.[49] However, if two vehicles (an autonomous and a human-driven one) collide, other provisions of the HCC are to be applied. At first, the one whose behaviour was at fault (i.e. the damage is attributable to him or her) is liable to provide compensation.[50] It is remarkable that crash reports show the frequent occurrence of rear-end traffic collisions. One of the reasons of these accidents is the extremely short reaction time of the AVs for staying in the lane and using the breaks in ways for which the slightly undisciplined human drivers cannot be properly prepared. Under the above-mentioned provision, the human driver (and the insurance company) is the aggrieved party who has to bear all cost consequences in these cases as an autonomous car cannot act with fault. Therefore, in our opinion, the application of this provision will not be appropriate for these collision cases.

47 http://www.europarl.europa.eu/RegData/etudes/STUD/2018/615635/EPRS_STU(2018)615635_EN.pdf.

48 Nyholm–Smids 2018.

49 HCC Sec 6:535 Subs. 3.

50 [Interaction of hazardous operation and relationship of operators in liability for torts committed jointly]

(1) Where damage is caused by one hazardous operation to another, the operators shall be liable to provide compensation as commensurate according to attributability. If the damage is caused by a person other than the operator, the operator shall be liable to provide compensation as commensurate according to the attributability of the de facto tortfeasor.

(2) If the cause of damage is not attributable to either party, compensation shall be provided by the party whose dangerous activity is responsible for the malfunction that contributed to causing the damage.

(3) If the cause of mutual damage is a malfunction that occurred in the scope of both parties' dangerous activity, or if such malfunction cannot be attributed to one of the parties, each party shall, where individual responsibility cannot be established, bear liability for their own loss.

5.2. The Capacity of the Operator of Dangerous Activity

The capacity of an operator of a hazardous activity (the operation of an AI system) determines the human who must carry the risk in case the AI causes damage, regardless of its (or the driver's) personal conduct or blameworthiness.

In the following, the concept of the keeper of hazardous things should be examined more accurately as being the operator who carries out the hazardous activity or the person who assumes control over that activity. However, it should be mentioned here that administrative rules also affect the civil law concept of keeper through the obligation for registration or for concluding third-party insurance contracts.

Within the scope of the meaning of *operator*, the HCC of 2013 brings some novelty. The judicial practice developed and fixed the concept of the operator decades ago under the old HCC of 1959. The operator is the person who maintains and continuously undertakes the hazardous operations or under whose oversight management and/or control of the hazardous operation would be undertaken.[51]

The use of an autonomous car would be a good example for choosing a scenario for examining the concept of operator and for establishing the details of who will be held liable. At this time, the keeper of a motor vehicle is quite fixed and determined. It is obvious that there is an inconsistency between the registered person and the person who controls the vehicle.[52] In the case of vehicles, the operator is the one who has actual and economic control over the vehicle but who is not necessarily its legal owner.[53]

According to the Ethical Guidelines of EU-HLEG:

> human oversight may be achieved through governance mechanisms such as a human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC) approach. HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. HIC refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation.[54]

51 Old decision of Supreme Court of Hungary No. BH1988. 273. The decision added that the fact that who has the interest in the ultrahazardous operation is irrelevant. This decision needs to be revised in the light of the new HCC provision.

52 For the discrepancies between the administrative and the civil law concept, see: Pusztahelyi 2018c. 216–229.

53 Karner 2018. 370.

54 HLEG: Ethics Guidelines for Trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, 16.

It means that the human oversight would decrease more and more upon the artificial intelligent system as AI technology develops gradually and as general AI will be rolled out. For example, in the case of autonomous cars with a maximum automation, the AV operates and fulfils tasks without any expectation that a user will respond to a request to intervene. On levels 4 or 5 of driving automation, the user's role is only to verify operational readiness of the ADS-equipped vehicle and to determine whether to engage the ADS. The user becomes a passenger only when the ADS is engaged if physically present in the vehicle.[55]

The actual control of the operator will fade, so the classical assessment of the capacity of the operator needs to be revised. Finally, we should stress that the legal concept of operator (keeper) is a stringent and mandatory rule. It means that the agreement to shift the capacity of operator to someone else is invalid if the personal and technical requirements to take over control are missing. This rule does not affect the validity of a contract which is concluded for undertaking the losses emerging from this case. The new rule of the HCC emphasizes only one element of the concept of operator: 'The person *on whose behalf* the hazardous operation is carried out shall be recognized as the pursuer of a dangerous activity'[56] (emphasis added).

This change would be assessed as a shift towards *the principle of aktive interesse (who is the person gaining benefits from the operation of the hazardous activity*) and in our opinion towards the risk-allocating function of the liability rule *and the risk management approach.* According to judicial practice, the user who uses the AI system for their own purposes (for example, who engages the motor and switches on the autonomous system) will be deemed as being an operator and will be held liable for damage caused by an AI system (i.e. autonomous car). However, this person is not able to control and to correct the malfunctions of the autonomous system; moreover, the operation of the AI system is hardly understandable, untraceable, or uncontrollable for him or her, particularly in the case of the so-called black-box phenomena. In our opinion, at this point of development, the concept of operator within the above-mentioned classical scope of meaning will begin to become inappropriate for application relating to AI technology. Therefore, when one of these above-mentioned new types of risks[57] manifests itself as the cause of the damage, the operator has got less opportunity to take protective measures or act promptly.

In the scope of these new risks, in most cases, the manufacturer is the person who can manage the risk and has the means to defend against them. In this

55 SAE: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles J3016 of Jun 2018. 21.

56 HCC Sec 6:536. Subs. 1.

57 Risks relating to the failure of the operating software that enables the AVs to function, risks relating to network failures, risks relating to hacking and cybercrime, and risks/externalities relating to programming choice.

viewpoint, the manufacturer can be deemed as the operator. We see that the function of societal distribution should be the focus of the solution given, but we experience meanwhile some departure from the other rationale (on whose behalf the dangerous activity is carried out). The two different approaches for the capacity of operator lead to the uncertainty of the legal concept and lead us to examine the manufacturer's role as operator of the AI systems. The question is whether the liability rule for dangerous activities is appropriate in order to establish the accountability of the manufacturer or whether these problems need an innovative approach to liability rules. In the cases when the control shifts to the manufacturer, this person will be the operator of the AI device. As a result: 1) the scope of the activity would be largely extended – as I mentioned – and 2) the evitability of the external causes leading to damage would lose its subjective side (i.e. the possibility to take action promptly when the AI device itself is not able to cope with it). The manufacturer is far removed from the accident scene and is unable to influence the behaviour of the vehicle in the relevant situation.

One more opportunity should be examined in order to retain the operator's liability. In theory, there will be a possibility to separate the human's failure and that of the AI with the help of a black-box recording device. *The human or non-human agent who actually takes control of the machine could be considered the operator of the hazardous activity, who could technically change from time to time.* It is possible to share the capacity of the operator between the user (driver, owner) and the manufacturer. *But the concept of operator requires also some stability and continuity.* Relating to one single hazardous operation, the capacity of the operator is singular, with the exception when several persons' interests are common. This concept excludes that another person can be held liable for any malfunction at the same time and grounded on the same liability.

The above-mentioned study of the European Parliamentary Research Service states that *the application of the existing rules to AVs* will likely shift the existing balance *in liability distribution between consumers and producers, further accentuating existing gaps and potentially contributing to legal and administrative costs arising from uncertainty.*[58] Gerhard Wagner states the following: 'The risk management approach, envisaged to serve as an alternative to strict liability, should not, it is said, focus on the person who acted negligently but rather on the individual who was able to minimize risks and deal with negative impacts.'[59]

These are the reasons why we agree with Gerhard Wagner, who emphasizes the shift from user control to manufacturer control, relating to the liability for dangerous activity.

58 http://www.europarl.europa.eu/RegData/etudes/STUD/2018/615635/EPRS_STU(2018)615635_EN.pdf, 24.

59 Wagner 2019. 30.

6. The Future of the Strict Liability for AI Technology

It should be mentioned that liability for dangerous activity is interpreted nowadays as the ceiling of the standard of reasonable care as the general non-contractual liability rule is also based on fault with objective meaning. Nevertheless, the objective approach to strict liability rule for dangerous activity may trigger two consequences. First, the characteristics of the liability rule will change as the blameworthiness will fade out and the liability rule will be degraded to a pure risk management solution, where the deterrence function disappears and only the societal distributive function remains. Second, in parallel with the fact that the elements of a special case of liability (especially the dangerous nature, the causal link between hazardous operation and damage) will weaken and the scope of the application of this liability rule will broaden, the scope of the general fault-based liability rule will diminish. Therefore, the strict liability rule for (not so) dangerous activities will compete not only with the other strict liability rules (e.g. product liability) but also with the general fault-based liability rule.

In cases when the manufacturers will be held liable on the grounds of this strict liability rule, as operators, product liability and the liability for hazardous operations will come very close. Their conceptual scopes will graze each other, and this fact can degrade the priority of product liability which the EU legislation and CJEU practice want to strengthen.

We think that this is a useful liability rule which served for rendering good solutions and as a legal basis for indemnification relating to disturbances of technological development for more than one century. At the roll-out of autonomous systems, as we stated above, interpretational and application questions will emerge relating to this rule.

To sum up, in the phase when general AI will be rolled out and will appear in the retail sector, the strict liability rule for dangerous activity will begin to lose its potential relating to general AI applications. Although this strict liability is worded in a general clause which generates fewer administrative costs and allows the injured person to gain compensation in a relatively easy way, the capacity of the operator – with its original conceptual scope – will be insufficient at the following stage of AI development.

References

BAKER, T. 2005. Liability Insurance as Tort Regulation: Six Ways That Liability Insurance Shapes Tort Law in Action. *Connecticut Insurance Law Journal* 12(1): 1–16.

BATTESINI, E. 2005. Tort Law and Economic Development: Strict Liability in Legal Practice. *The Latin American and Iberian Journal of Law and Economics* 1: 7–9.

BORGHETTI, J. 2019. How Can Artificial Intelligence Be Defective? In: *Liability for Artificial Intelligence and the Internet of Things: Münster Colloquia on EU Law and the Digital Economy.* Baden-Baden. 63–76.

DE GRAAF, R. 2017. Concurrent Claims in Contract and Tort: A Comparative Perspective. *European Review of Private Law* 4: 701–726.

DEZSŐ, Gy. 1932. A kártérítési kötelezettség különféle alapjairól. In*: Glossza Grosschmid Béni.* Fejezetek kötelmi jogunk köréből *című művéhez*. Budapest.

EHRENZWEIG, A. A. 1966. Negligence without Fault. *California Law Review* 4: 1422–1477.

EÖRSI, Gy. 1972. A szerződésen kívüli felelősség és a Ptk. reformja. *Jogtudományi Közlöny* 3: 65–75.

1975. The Validity of Clauses Excluding or Limiting Liability. *The American Journal of Comparative Law* 2: 215–235.

FAZEKAS, J. 2017. A kontraktuális és deliktuális felelősség viszonya az új Polgári törvénykönyvben, különös tekintettel a párhuzamos igényérvényesítést kizáró non-cumul szabályra. In: *Állandóság és változás. Tanulmányok a magánjogi felelősség köréből*. Budapest. 24–52.

FIORE, K. 2017. No-Fault Systems. In: *Encyclopedia on Law and Economics*. Elgar Online.

FUGLINSZKY, Á. 2017. Some Structural Questions on the Relationship between Contractual and Extracontractual Liability in the New Hungarian Civil Code. In: *New Civil Codes in Hungary and Romania*. Cham.

GRISS, I.–KATHREIN, G.–KOZIOL, H. (eds). 2006. *Entwurf eines neuen österreichischen Schadenersatzrechts*. Vienna.

GROSSCHMID, B. 1900. *Fejezetek kötelmi jogunk köréből*. Budapest.

HUGUENIN, C.–HILTY, R. M. (eds). 2013. *Schweizer Obligationenrecht 2020: OR 2020: Entwurf für einen neuen allgemeinen Teil*. Zürich.

JUHÁSZ, Á. 2013. The Regulatory Framework and Models of Self-Driving Cars. *Zbornik Radova Pravnog Fakulteta u Novum Sadu* 3: 1371–1389.

KARNER, E. 2018. A Comparative Analysis of Traffic Accident Systems. *Wake Forest Law Review* 53: 365.

KOLOSVÁRY, B. 1908. Automobiljog, különös tekintettel az új osztrák automobiltörvényre. *Erdélyrészi Jogi Közlöny* 37–38: 368–377.

LÁBADY, T. 2014. [Commentary of Section 6:535]. In: *Kommentár a Polgári Törvénykönyvhöz*. Vol. II. Budapest.

MARTIN-CASALS, M. (ed.). 2014. *The Development of Liability in Relation to Technological Change*. Cambridge.

MARTON, G. 1931. Veszélyes üzem. *Polgári Jog* 4: 147–161, 6: 242–255.

NYHOLM, S.–SMIDS, J. 2018. Automated Cars Meet Human Drivers: Responsible Human-Robot Coordination and the Ethics of Mixed Traffic. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-018-9445-9.

PUSZTAHELYI, R. 2016. Igényhalmazatok a szerződésszegési jogkövetkezmények rendszerében, különös tekintettel a Ptk. 6:145.§-ára. *Pro Futuro* 2: 60–78.

2018a. Veszélyes üzemi felelősség a Ptk. tükrében (conference paper, manuscript; Az új Polgári Törvénykönyv első öt éve. MTA TK JTI – ELTE).

2018b. Veszélyes üzemi felelősség: az objektív kártérítési felelősség térnyerése. *Gazdaság és Jog* 9: 3–8.

2018c. Az üzembentartó polgári jogi fogalma és a közigazgatási bírság „címzettje" – gépjárművek üzemben tartása. *Pro Publico Bono* 3: 216–229. https://folyoiratok.uni-nke.hu/document/nkeszolgaltato-uni-nke-hu/WEB-PPB_2018_3---216-229_Pusztahelyi.pdf.

REISCHAUER, R.–SPIELBÜCHLER, K.–WELSER, R. 2006. Reform des Schadenersatzsrechts. Vorschläge eines Arbeitskreises. Vienna: Ludwig Boltzman Institute–Manz Verlag.

RICHARDS, N. M.–SMART, W. D. 2016. How Should the Law Think about Robots? In: Calo, Ryan–Froomkin, A. Michael–Kerr, Ian (eds), *Robot Law*.

VAN DAM, C. 2013. European Tort Law. Oxford University Press, 2nd edition.

WAGNER, G. 2006. Tort Law and Liability Insurance. *The Geneva Papers on Risk and Insurance. Issues and Practice* 31(2) – *Special Issue on Law and Economics and International Liability Regimes* (April): 277–299.

2019. Robot Liability. In: *Liability for Artificial Intelligence and the Internet of Things. Münster Colloquia on EU Law and the Digital Economy IV*. Baden-Baden: Nomos. 27–62.

Acta Univ. Sapientiae, Legal Studies, **8**, 2 (2019) 231–244

DOI: 10.47745/AUSLEG.2019.8.2.06

Lawyers and the Machine. Contemplating the Future of Litigation in the Age of AI

János Székely
PhD, Senior Lecturer
Sapientia Hungarian University of Transylvania, Cluj-Napoca
Department of Law
E-mail: szekely.janos@kv.sapientia.ro

Abstract. The possible impacts of artificial intelligence (AI) on the modern world constitute a complex field of study. In our analysis, we attempt to explore some possible consequences of the utilization of AI in the judicial field both as regarding adjudication, formerly exclusively reserved for human judges, and in the rendering of legal services by attorneys-at-law. We list the main factors influencing technology adoption and analyse the possible paths the automated management and solution of disputes may take. We conclude that the optimal outcome would be a cooperation of human and artificially intelligent factors. We also outline the conditions in which, following the abandonment of the principle of procedural fairness, AI may be directly utilized in judicial procedure. We conclude that big data solutions, such as social rating systems, are particularly concerning as they constitute a conceivable modality of deploying AI to solve litigious disputes without regard to fundamental human rights as understood today.

Keywords: artificial intelligence, automatic decision making, attorney-at-law, judges, social rating system

1. Introductory Thoughts

The possible impacts of artificial intelligence (AI) on the modern world constitute an ever more complex field of study. Speculations abound regarding the effects, both benign and malign, which developments in this field may have in the world of work, business, education, the public and the private sphere. There are already tangible implementations of AI but far fewer than the proposed uses. As AI is likely to touch all fields and domains of human activity, even if the stark warnings of some detractors are unlikely to materialize, we must proactively contemplate

its effects. In our study, we attempt to explore some possible consequences of the utilization of AI in the judicial field, both as regarding adjudication, formerly exclusively reserved for human judges and other similar personnel, and in the rendering of legal services, by attorneys-at-law. Lawyers – in the wider sense of the term (referring to all experts of law, regardless of their profession) – will inevitably be affected by the emerging uses of AI. Some authors[1] have explored this question with a varying degree of optimism, pessimism, and sense of certainty about the changes which may occur, prophesizing both upheaval and gradual adaptation. Such predictions should, however, be carefully scrutinized.

2. AI and the Adoption of New Ideas

All change, be it economic, technological, or – alas – even legal, may only take place if a problem, a more optimal solution than those previously available,[2] the political will for implementing such a solution, and a popular desire to have the solution implemented are available, all at the same time. History abounds with examples of solvable problems which remained unsolved even though the concepts, means (such as inventions), and methods (such as legal norms) meant to resolve them were already available. We need not look further than one of the oldest, gravest problems which 'plagued' mankind: disease. Microscopes were available as early as the 17th century, and with them also the knowledge of microorganisms. The possibility that these so-called 'animalcules' may cause disease was raised simultaneously with the advent of microscopy.[3] Yet it was only in the late 19th century that germ theory became accepted as scientific fact, leading to the employment of pre-existing means for treating a pre-existing problem. Now defunct theories of transmissible disease, rooted in irrational notions inherited from antiquity, such as the miasma theory,[4] lingered on long into modernity, a supposedly more rational age, without any scientific evidence to support them. For lack of popular acceptance of a scientific solution, countless lives were lost.

We mean not to digress here but to provide a useful analogy to which we may refer to in the following analysis of the proposed effects of AI on the legal professions. The fact that germ theory failed to 'catch on' for several centuries, even in the light of mounting evidence, should caution us whenever we contemplate the usefulness of discovery, or scientific and technological innovation for solving problems, even when in theory such innovation would be game-changing. This is all the truer in the legal field, strongly permeated by a rich mesh of intertwined

1 See Susskind–Susskind 2015.
2 See Kuhn 1970.
3 Williamson 1955. 46.
4 Williamson 1955. 45.

interests, tradition, institutions, and politics. After all, the personal computing revolution has been ongoing for decades, yet the usefulness of computers as veritable replacements for the human factor in the justice system is only now being seriously contemplated.

The complexities of litigation (familiar to attorneys) and the intricacies of adjudication, which sometimes challenge the best and brightest human judges, are all too well known and should not be reiterated here. When thinking about the implementation of AI to automate these processes, we should not forget that the much simpler activities routinely undertaken by other legal professionals, such as public notaries, have not yet been automated. While secure authentication of persons is possible (inter alia, by use of various forms of biometric information), rendering contracts concluded in electronic form all but irrefutable (in the same way banking operations conducted over the Internet are considered to be), the cooperation of a notary public is still required by law for the validity of certain deeds in many countries, even when these are no more complex than filling out forms with predetermined contents and then signing them. A computer system is as able as any human being to ascertain the identity of the signatory, the fact that the document has been filled out correctly as well as the date at which it was concluded. Such systems have been able to do so for nearly two decades, yet notaries public did not and do not seem to be threatened by the kind of 'transformation' akin to extinction Susskind and others envision for attorneys-at-law in their current form. So, the question arises: will AI ever even be implemented in the judicial field?

In order for us to even attempt an answer to this question, we must, even if superficially, delve into the dizzying array of technology adoption models which have been developed over the years.[5] Various models offer various answers to the factors which most influence technology adoption, but some common traits can be discerned from these. In more recent research, the so-called Unified Theory of Acceptance and Use of Technology (UTAUT) model has been developed in order to predict the adoption of new technologies.[6] This model emphasizes the importance of *behavioural intention*, that is, the intention of a person or organization in adopting new technology. While the UTAUT model is vastly more complex than may be presented here, the volitional element of behavioural intention should be emphasized for the purposes of this study.

This intention is augmented by the belief that utilization of the system will increase performance or productivity (*performance expectancy*). If the potential users believe the new technology to be easy to use, this will also count towards its adoption. The opposite is true if the technology is expected to be difficult to use (*effort expectancy*). Lastly, if it is believed that the institutional framework

5 See Patel–Connolly 2007.

6 Viswanath–Morris–Davis–Davis 2003. 446–467.

which offers support in the use of new technology (such as easy-to-access advice, training, etc.) is present, this will also contribute to its adoption, while the absence of such framework will discourage the adoption of new technology (*facilitating conditions*). The demographic profile of the users (age, gender), their experience (i.e. technological experience), and the degree to which some have already voluntarily adopted the new technology may also count for, or against, its wider spread.

Taking these factors into account separately, and in various particular situations, will be the key to predicting whether AI will ultimately 'catch on'.

3. The World of Tomorrow in the Judicial Process

The conclusions of the Woolf report regarding civil justice (and perhaps justice in general) are all too familiar to us, even decades after they were first put to paper: 'The key problems facing civil justice today are cost, delay and complexity.'[7] Here then lies the problem to be solved.

The means to solve it, information technology, has been with us for nearly three decades. Yet the solution seems not to have been applied to the problem. Even in developed jurisdictions, not to speak of Eastern Europe, solutions based on information technology cannot be considered abundant, with the best intentions (as is evident from other writings in the present issue of this journal).

A patchwork of experimental schemes and pilot projects cannot reasonably be deemed a revolution, yet the predictions of Susskind and others are unwavering: a new era is upon us, when technology will – eventually – transform the legal profession. We mean not to say that technology has not brought any change at all: the activity of attorneys-at-law and that of judges was to a certain degree transformed by the use of computers, e-mail, real-time image and voice transmission, the ready availability of searchable legal texts and of jurisprudence. All these may be deemed a progress in themselves. However, the predicted revolution failed as of yet to materialize. Computer technology was never extended into the courtroom and into the mind of the adjudicator itself in an all-encompassing manner. The attorney or the judge may have access to electronic resources, to the case file in scanned form, even with searchable content, to the applicable law, and to the relevant jurisprudence in electronic databases. Yet weighing the facts, applying the law, upholding procedural guarantees, and rendering the decision have not yet been automated. As with other emerging technologies, such as the blockchain,[8] and the age of cryptocurrencies and the smart contracts it heralds, AI seems to have delivered a lot less then promised.

7 Wolf 1997. 709.

8 For a few such predictions, see: Flood–Robb 2019.

This apparent failure is due to the ways in which the implementation of emerging technology tends to unfold but also to the quite limited capabilities of the technological solutions themselves. It is not just necessary for a technology to be possible or even available. A myriad of factors influences its percolation. Performance expectancy, effort expectancy, and facilitating conditions, intrinsically linked to the institutional framework in which the new technology is to be deployed, all have their roles to play.

4. What Can AI Do for Lawyers

Whenever we think of AI, the concept of artificial intellect, or artificial general intelligence (AGI) comes to mind.[9] An intelligent and perhaps omniscient entity, capable of perceiving the material world, understanding spoken and written human language, cognition and emotion, of rendering complex judgements with utmost speed and objectivity is still the technology of tomorrow, and it is possible it will always be. Too many predictions of the future are still based on this utopian concept. The methods for attaining an ever more generalized form of artificial intelligence are numerous and diverse. For non-initiates, Boden lists these simply as 'heuristics, planning, mathematical simplification, and knowledge representation'[10] methods.[11] We shall not attempt to present these methods here, limiting ourselves to stating that the field of AI is a populous zoo filled with all manners of creatures, having sometimes wildly differing characteristics. Therefore, AI is an insufficiently precise concept when dealing with its applications, both current and future, in the judicial field.

The form of AI most often referred to in discussions nowadays is called machine learning[12] (although this concept is only marginally less fuzzy than artificial intelligence itself). This is limited to discerning or recognizing pre-existing patterns in large amounts of data and offering a certain output based on the patterns recognized. The methods used for pattern recognition may vary, making them a universe onto itself, the functional intricacies of which are better left to studies of a more technical nature. What should be emphasized here are the effects of such machine learning algorithms, specifically their uncanny ability for pattern recognition in apparently unrelated data and for prediction of apparently inscrutable future outcomes. It is our view that these effects should be the main focus of study when the impact of artificial intelligence on legal professions is examined.

9 See Boden 2018. 18–19.

10 See Boden 2018. 20–49.

11 For a discussion on AI methods as applicable to the activities of a judge, see also: Schubbach 2019.

12 See: Boden 2018. 69–89, Johnson 2019. 1232–1239.

Of course, machine learning is quite apt at data systematization and retrieval too, which may also benefit the judicial process by eliminating the human factor so relevant in obtaining, processing, and presenting evidence and in working out and refining the legal argumentation in the case. These aspects of AI, however (while impacting lower and even higher-added-value legal work, as correctly recognized by Susskind), offer little in the way of revolutionary change, simply constituting a more evolved form of what expert systems were meant to do. These were developed beginning all the way back in the 1980s and were intended to achieve a limited goal: an automated way of assisting human experts, complementing their abilities, by removing non-creative repetitive tasks from their workload.

Since developments in the field of information gathering, systematization, and retrieval now permit a wider deployment of such systems, their use has mushroomed. Thanks to machine learning, they are now being used to identify relevant judicial precedents, sort the 'wheat from the chaff', during litigation by filtering documentary evidence to discern the admissible from the inadmissible,[13] and so on. They make the work of attorneys and judges easier but tend not to replace these professions, only to augment their abilities. In this view, lawyers and the machine may coexist in a feedback loop in which big data systems permit human operators to better document cases in fact and law, leading to better decisions which in turn result in a more constant jurisprudence, which feeds back into databases for such jurisprudence, parsed by AI and presented to human operators in order to further refine legal argumentation, and so on. In such implementations, human beings are – by definition – in the loop.

In this model of thought, the implementation of AI by lawyers, and in the judiciary as a whole, should be imminent and inevitable as all the requirements for the adoption of new technology, as presented above, would be conducive to such a result. In this feedback loop, the intrinsic understanding of the technology employed is almost unnecessary so long as it yields satisfactory results to human attorneys, which are found acceptable by human judges. Thereby, a high performance expectancy and a low effort expectancy would be associated to these solutions. They would also be facilitated by the need for an efficient, low-cost judicial system.

Since the machine does not take over decision making from the human factor, the world would be considered unchanged when the societal and political dimensions of rendering judicial decisions, a phenomenon masterfully described by Damaška,[14] are concerned. The judge would be free to consider the legal reasoning when solving the dispute, while leaving room in the decision for the implementation of whatever policies the character of judicial (civil or criminal) procedure is meant to convey (as all procedural systems in Damaška's view do to

13 See Keeling–Huber-Fliflet–Jianping–Chhatwal 2019.
14 See Damaška 1986. 147–180.

some extent). This would be the 'better mousetrap' view of artificial intelligence when applied to lawyers: a mechanism (however complex) for attaining improved outcomes in an institutional system which remains unaltered when it comes to the fundamentals of its workings.

This future of AI should not concern us any further since it would not alter the framework in which the activity of legal professions takes place. By not delegating decision making to an algorithm, merely using it to automate the information gathering phase of the procedure (the collection, sorting and indexing of evidence, the identification of applicable law and judicial practice, or precedent when necessary in the given legal system), we can assure respect for procedural guarantees and prevent bias. Such a model for AI implementation would also alleviate the issues of opacity[15] and the lack of human-readable reasoning, which necessarily arise if we adopt the model of a robot judge, or automated litigation, as we shall see in the following. We deem this modality of AI adoption to be preferable to all others when the future of the legal professions is concerned.

5. ADM – The Robot Judge

As opposed to an AI-assisted future of judicial procedure, in which computers are relegated to providing the ingredients to a well-founded decision, there lies the model of *automated adjudication* (automated claims processing). In this model, AI would not (only) be engaged in gathering the necessary information for assisting human beings in rendering a judicial decision but would also either propose or, indeed, impose the contents of such decisions.

There have been attempts with varying degrees of success in implementing systems for automated claims processing. For these, various orders for payment or small claims procedures in different countries may be provided as examples. However, in these cases, automation does not apply to any judgements on the merits of the claim but simply to the automated management of the creation, storage, postage, and, if applicable, enforcement of documents, which can scarcely be called judicial decisions. They are, in reality, documents attesting to a debt which may be enforced if the debtor was not diligent enough or lacked the ability to contest their contents in the time period provided. They are, in essence, no different to invoices issued automatically by electronic billing systems. No adjudication activity takes place prior to them being issued, and no procedural guarantees, such as a right to defence, are provided. Any such guarantees are reserved for the judicial procedure, which might take place if the decisions rendered are contested. Such systems present significance from the standpoint of AI because they show a desire by legislators to replace the human judge or,

15 See Chesterman 2020.

better said, adjudication as an activity, whenever this is perceived as feasible. The tendencies for the implementation of such systems also show the possible places in which we should look for the first true AI judges. Summary procedures or small claims procedures tend to be predictors of the direction in which procedural norms are likely to develop in the future, and this may be the case also when AI implementations are concerned. Voluntary procedures specific to the sphere of alternative dispute resolution (ADR) should also be watched as they are more likely than not to become testbeds for AI technology in adjudication due to their confidential nature and the much laxer procedural guarantees applicable to them.

An interesting attempt at implementing a true AI judge is underway in Estonia[16] for claims not exceeding 7,000 euros; however, its results are not yet widely visible. This mode of litigation is not similar to automatically issuing small claims decisions as the AI agent would in fact act as a true judge, analysing submitted documentary evidence on its merits and rendering a solution, which would only be subject to appeal to a human judge. The meaning of this 'appeal' is not yet known. But as anyone versed in judicial procedure knows, the notion of appeal may hide varying degrees of judicial review: it may refer to full review (of the facts, the substantive law, and applied procedural norms, or, as the case may be, of judicial precedent) but also to partial review (where only judicial errors of a certain type or gravity are analysed). Appeals may also be subjected to formal requirements, such as legal representation, and may presuppose the advance payment of a fee or tax prior to being considered. All these factors added to a likely submission by judges to the decisions considered issued by a superior entity may in turn erode procedural protection. This type of AI implementation in judicial procedure is called automatic decision making (ADM).[17]

We should not make the mistake of thinking that, once deployed, AI judges will remain exiled to the realms of small claims or even of civil procedure. Already, AI is used to predict the risk of recidivism in criminal procedure, where it shows a concerning degree of bias, due to the data utilized for feeding, or teaching the algorithms.[18] The algorithms which such applications are based on are already deployed in the private sector in medical implementations, in the labour market, in the financial sector, and mostly anywhere where their powerful predictive abilities can be harnessed.[19] AI's ability to predict judicial decisions, for example, in cases when human rights are at stake,[20] is particularly concerning. The risk of bias in implementations of AI is a topic readily discussed in the scientific literature.[21]

16 Niiler 2017.
17 Johnson 2019. 1219.
18 For a discussion of AI use in the COMPAS (Correctional Offender Management Profiling for Alternative) system for predicting recidivism see: Chesterman 2020. 3–6.
19 Johnson 2019. 1215–1217.
20 Lu 2019.
21 See Johnson 2019. 1239–1245. For a detailed albeit technical description of bias in data-driven systems, see: Ntoutsi et al. 2020, Howard–Borenstein 2018.

We must underscore here the types of implementation that an AI judge is currently considered to be applicable to: rendering decisions based on predictions, predictions which in turn result from massive amounts of data that were processed in order to discover correlations (importantly, not causation) which cannot be easily perceived by the human intellect.[22] The AI judge, as things stand, is able to solve cases by predicting behaviours,[23] which are set to take place in the future, or to retrospectively determine which would have been the most likely course of action taken in the past. The quality of such predictions improves as time goes by; however, they remain *predictions*, based on abstract assumptions, not at all time grounded in the realities of the particular cases they are being applied to but founded on the aggregation of big data[24] knowledge.

6. Why One Should Dread the Robot Judge

This predictive, mostly deductive nature of ADM technologies raises the spectre of decision-making mechanisms quite unlike those we are currently used to, which will be applied in the judicial procedure. The crux of the problem here is that the inherently opaque nature of AI, as discussed by Chesterman,[25] is quite incompatible with the desired qualities of a fair trial. In fact, if the notion and prerequisites of a fair trial – as outlined by the European Court of Human Rights, for instance – would remain unchanged, ADM should be considered completely contrary to such a notion, and inadmissible. This would limit the application of ADM mechanisms in judicial procedures to alternative dispute resolution, where, in the measure permitted by law, the free will of the parties prevails over procedural fairness.

The methods by which ADM would be conducted are often inscrutable to human beings.[26] Even if we assume the best intentions of the constructor of such methods, believing that this opacity is not meant to conceal malicious intent or inadvertent bias, the fact remains that ADM mechanisms are unable to give reasons for their decisions in a human-readable, intelligible form. In order to comply with the requirements of a fair trial, a court must be able to determine and then describe in a human-intelligible way the factors on which its decisions are based.[27] Only in this manner may the fairness of some elements of procedure, such as impartiality

22 For a detailed recent analysis of this issue, see: Chesterman 2020.
23 Johnson 2019. 1232 et seq.
24 Ntoutsi et al. 2020. 4.
25 See Chesterman 2020.
26 Chesterman 2020. 4–8, Schubbach 2019.
27 For a description of the jurisprudence of the European Court of Human Rights which imposes this requirement, see: Guide on Article 6 of the European Convention on Human Rights – Right to a Fair Trial (civil limb), 71–72.

and independence (lack of bias) of the court (or the AI judge for that matter), be fully assessed.[28] Also, the right to appeal, if provided by the given procedural system, cannot be exercised should the reasons for the decision being appealed be inscrutable to the appealing party. During any appeal, the requirement for a fair trial must also be, in principle, observed.[29] It is questionable if an AI entity may even be considered a tribunal[30] in the meaning of Article 6, paragraph 1 of the European Convention on Human Rights as the text was drafted in an age when the rendering of judgements by an algorithm was unfathomable.

If we accept that ADM will become part of the procedural landscape of the future, we must unshackle ourselves from basic notions of procedural fairness, taken for granted today. In fact, if the fairness of ADM procedures is to be evaluated, such evaluations are likely to be performed by non-lawyers turning a field already thought to be technical to one quite unintelligible to non-initiates. We consider that the facilitating conditions for such AI solutions as the 'robot judge' are not yet present in Western democracies which remain beholden to the notions of fundamental human rights and, among them, to procedural fairness or the due process of law.

Since the era of the Woolf Report,[31] the tendency in civil procedure has been to make litigation more accessible to the public while simplifying the process. This tendency for accessibility and simplification has seen the role of attorneys-at-law diminish, and an outright hostility against compulsory legal representation form, as a manifestation of efforts directed towards the democratization[32] of justice. For the past decades, the tendency has been to make the application of law less technical, and especially to reduce the role of attorneys[33] (the legal 'profession' to which Susskind most often refers to) in the judicial process. If we accept this tendency as a righteous one, aimed at improving access to justice for the poor, the disadvantaged, and those being discriminated against, then, surely, the solution to their plight cannot be constituted of making a system which still remains labyrinthine even less intelligible to non-professionals. We should add to this the intrinsic incompatibilities of ADM with basic principles of a fair trial, themselves constituting elements of a fundamental human right, while the opposition to ADM is likely to be significant as well. Therefore, we deem it unlikely that ADM would gain prevalence based on popular demand, so long as current trends hold. Lack

28 See: Guide on Article 6 of the European Convention on Human Rights – Right to a Fair Trial (civil limb), 44–53.

29 See: Guide on Article 6 of the European Convention on Human Rights – Right to a Fair Trial (civil limb), 16–18.

30 See: Guide on Article 6 of the European Convention on Human Rights – Right to a Fair Trial (civil limb), 33–34.

31 See Woolf 1995.

32 Assy 2015. 15–21.

33 Backer 2018. 128.

of popular demand, or acceptance for a new technology, may hinder its adoption – as we have seen – even if the technology itself is available and the problem it is meant to solve is widely known.

7. Conclusions. ADM and Current Dangerous Trends

Of course, as Cohen notes,[34] it is possible that judicial procedure in itself will continue to fragment along already existing fracture lines and consolidate widely differing regimes for differing types of litigation, removing certain types of claims from the court process altogether, thereby exposing them to alternative dispute resolution schemes. Such schemes may be more prone to the implementation of ADM. Also, this possible future might result in the 'balkanization' of judicial procedure, yielding various procedural regimes, some based on ADM, others on a human judge.

Popular demand may not be the only driver for the adoption of new technologies. ADM may not just be proposed to ease access to the judicial system, by making procedures faster and cheaper, but it may also be imposed, over the heads of stakeholders, by a state or other entity interested in cost efficiency, reducing the risk of corruption or extending state authority into the judicial process by transforming it into a vehicle of policy implementation as is known to happen[35] in authoritarian regimes. This would, as a necessity, presuppose the transformation of the meaning of a fair trial and that of the notion of judge, and even the notion of justice itself. If we are ready to abandon such requirements and accept the rendering of judicial decisions as a result of statistical probabilities determined by mechanisms inscrutable to most of us, then ADM and the AI judge may become a reality.

In judicial systems, where policy implementation is highly emphasized during the resolution of disputes, transition to ADM is all the more likely. If the structures designed for ensuring the rule of law, manifested in a respect for human rights, are subverted, systems for inducing social compliance, such as the social credit system set to be deployed in the People's Republic of China,[36] may emerge with significant effects in the field of adjudication.

> The Social Credit Initiative is a product of China's 'top-level design' (…) approach; coordinated by the Central Leading Small Group for Comprehensively Deepening Reforms. Its central objective is the development of a national reputation system: assigning a social credit number that reflects a qualitative

34 Cohen 2019. 154 et seq.
35 Damaška 1986. 8.
36 Backer 2018. 127.

> judgment of relevant data gathered about the subject. It will focus on four areas: 'sincerity in government affairs' (…), 'commercial sincerity' (…), 'societal sincerity' (…), and 'judicial credibility' (…). The term 'social credit' actually veils the overall character of the project. Sincerity in this sense means integrity and trustworthiness. The core object, of course, is built around the idea of compliance—that the way one complies with law and social obligation will be as important as the fact that one complies at all. That is a profound step forward from the more ancient forms of law and regulation. The former systems could be satisfied with the merest obeisance to its command; social credit systems judge compliance based on its effects given the spirit of the obligation or responsibility.[37]

If the judiciary is meant to primarily evaluate a tendency for compliance with existing norms, then it may be retooled in order to reward the likely more compliant party, while punishing the likely non-compliant party in any legal dispute. Effort expectancy for those implementing such a system may be low so long as they are not concerned with solving a legal dispute, only with disadvantaging one of the parties based on a perceived or predicted tendency to behave in a certain way, a tendency which may be evaluated, taking into consideration political views or other (such as social, cultural, or racial) factors. This behaviour has already emerged in AI, without any intention whatsoever. It should suffice to think of the COMPAS system and its biased actions against people predicted to be less compliant in the future. Social rating mechanisms have the added advantage of being palatable to the population, which sees in them the institutional manifestation of law and order expectations; thereby, it benefits from a high performance expectancy. Who would not want to live in a society where everyone respects the rules and refrains from antisocial behaviours?

In our view, the compliance enforcement model, in which big data is used to create an honesty rating, which is then utilized by an ADM agent (an AI judge), is much more likely to be adopted than any ADM solution which must meld the current requirements of a fair trial and the rule of law with the abilities of new technology. Efforts made in jurisdictions with legal systems which value compliance and collective action over individual rights may constitute a major facilitating factor for the adoption of such technologies. A social compliance rating system may even be 'sold' to the public as being 'democratic', given the ready acceptance of such rating systems already in use in social networking applications. The number of 'likes' one receives for one's posts on Facebook already incentivizes, or, for that matter, discourages behaviours of a certain type.[38] As correctly noted by

37 Backer 2018. 131.
38 Cohen 2019. 81.

Backer,[39] methods for social control employed by governments sometimes mimic those developed in the private sector (such as credit ratings), leading us from an age of collective rights to one of collective management.

No legal system should consider itself immune from this trend in which entire populations might be 'managed'[40] by a complex administrative framework – reliant on big data and artificial intelligence – of which the judiciary is only one component.

References

ASSY, R. 2015. *Injustice in Person: The Right to Self-Representation*. Oxford.

BACKER, L. C. 2018. Next Generation Law: Data-Driven Governance and Accountability-Based Regulatory Systems in the West, and Social Credit Regimes in China. *Southern California Interdisciplinary Law Journal* 1: 123–172.

BODEN, M. 2018. *Artificial Intelligence. A Very Short Introduction*.

CHESTERMAN, S. 2020. Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity. Forthcoming, A*merican Journal of Comparative Law; NUS Law Working Paper No. 2020/011*. https://ssrn.com/abstract=3575534 (accessed: 01.04.2020).

COHEN, J. E. 2019. *Between Truth and Power. The Legal Constructions of Informational Capitalism*. Oxford.

DAMAŠKA, M. 1986. *The Faces of Justice and State Authority: A Comparative Approach to the Legal Process*. New Haven (USA).

FLOOD, J.–ROBB, L. 2019. Professions and Expertise: How Machine Learning and Blockchain Are Redesigning the Landscape of Professional Knowledge and Organization. *University of Miami Law Review* 2: 443–482.

GUIDE on Article 6 of the European Convention on Human Rights – Right to a Fair Trial (civil limb). Updated to 31 August 2019 https://www.echr.coe.int/Documents/Guide_Art_6_ENG.pdf (accessed on: 2019.11.30).

HOWARD, A.–BORENSTEIN, J. 2018. The Ugly Truth about Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. *Science & Engineering Ethics* 5: 1521–1536.

JOHNSON, K. N. 2019. Automating the Risk of Bias. *George Washington Law Review Arguendo* 5: 1214–1271.

KEELING, R.–HUBER-FLIFLET, N.–JIANPING, Z.–CHHATWAL, R. P. 2019. Separating the Privileged Wheat from the Chaff – Using Text Analytics and Machine Learning to Protect Attorney–Client Privilege. *Richmond Journal of Law & Technology*. 3: 1–45.

39 Backer 2018. 140–149.

40 Backer 2018. 160–170.

KUHN, T. S. 1970. *The Structure of Scientific Revolutions*. 2nd edition. Chicago.
LU, D. 2019. AI Judges Make Good Calls on Human Rights Violations but Could Be Gamed. *New Scientist* 243 (17 August 2019): 8.
NIILER, E. 2017. Can AI Be a Fair Judge in Court? Estonia Thinks So. *Wired*. https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/ (accessed: 2019.08.07).
NTOUTSI, E. et al. 2020. Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey. *WIREs: Data Mining & Knowledge Discovery* 3: 1–14.
PATEL, H.–CONNOLLY, R. 2007. Factors Influencing Technology Adoption: A Review. *Information Management in the Networked Economy: Issues & Solutions*: 416–431.
SCHUBBACH, A. 2019. Judging Machines: Philosophical Aspects of Deep Learning. *Synthese*. https://doi.org/10.1007/s11229-019-02167-z (accessed: 2019.11.20).
SUSSKIND, R.–SUSSKIND, D. 2015. *The Future of the Professions: How Technology Will Transform the Work of Human Experts*. Oxford.
VENKATESH, V.–MORRIS, M. G.–DAVIS, G. B.–DAVIS, F. D. 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 3: 425–478.
WILLIAMSON, R. 1955. The Germ Theory of Disease. Neglected Precursors of Louis Pasteur. *Annals of Science* 1: 44–57.
WOOLF, H. 1995. The Woolf Report. *International Journal of Law and Information Technology* 2 (summer): 144–155.
1997. Civil Justice in the United Kingdom. *The American Journal of Comparative Law* 4 Symposium: Civil Procedure Reform in Comparative Context: 709–736.

Acta Univ. Sapientiae, Legal Studies, **8**, 2 (2019) 245–252

DOI: 10.47745/AUSLEG.2019.8.2.07

Artificial Intelligence and the Future of Labour Law

Dan Ţop
PhD, University Professor
Valahia University of Târgovişte, Târgovişte
President of the Association for the Study of Professional Labour Relations, Romania
E-mail: top.dan@gmail.com

Abstract. The notion of work goes through major changes caused by the development of technology, and it is assumed that the development of sophisticated robotization and artificial intelligence will undermine the existence of work. Artificial and robotic intelligence will create more jobs, not mass unemployment, as long as innovation is guided responsibly. Cobots, or collaborative robots, are typically intended for physical interaction with people in a common workplace. There is no doubt that the world of collaborative robots is on the rise, so labour law will have to distinguish between non-human workers (dwarfs, industrial robots, etc.) and human workers. Regulations in the field will evolve, meaning that provisions will be needed which will determine, at a minimum, what the relationship between the two categories of workers will be according to the specificity of the activity as well as other aspects. Romania still has a low density of 15 robots per 10,000 employees, with a national interest in the topic, which is a result of the adoption in 2015 of the National Strategy on the Digital Agenda Romania 2020. Replacing human labour with robots is no longer just a discussion, it is a reality; it is not just a sci-fi issue, it is something society should contemplate and anticipate by updating legislation and social protection.

Keywords: artificial intelligence, collaborative robots, digital inclusion, electronic person

Given that employers' needs are becoming more and more sophisticated, considering the new realities they face, the future of labour law can be more or less predictable. That is why we propose to look at one of the possibilities, not far from employers' predictable intentions, to resort more and more to the use of industrial robots. *Cobots*, or collaborative robots, are typically intended for physical interaction with people in a common workspace.[1] More and more

1 Colgate–Peshkin–Wannasuphoprasit 1996. 433–439.

industries seek[2] to replace employees with robots, which can work continuously and whose work is not taxed by the state.

The automotive and metalworking industries are the largest markets for robots, followed by electronics, plastic, food and pharmaceutical processing. These robots work alongside labourers and are flexible, easy to program, secure, and inexpensive. Romania had in 2016, according to the International Federation of Robotics,[3] 11 industrial robots per 10,000 industrial workers. A study by the World Economic Forum, with the title *The Future of Jobs*,[4] estimated that by 2020 more than 5 million jobs would disappear, affecting all industrial branches and all geographical regions. Although such dire predictions apparently failed to materialize, we are only at the beginning of robotization. Of course, any loss of jobs is predicted to be partially offset by the creation of new jobs in highly qualified fields.

The notion of work is undergoing major changes due to the development of technology, and it is assumed that the development of sophisticated robotization and artificial intelligence will undermine the existence of work. Artificial intelligence (AI)[5] and its impact on jobs have been important topics at the World Summit in Dubai in 2017. In a survey conducted by a US company, it is stated that about 65% of the children who are today in their first years of school will have jobs that have not yet been invented.[6]

However, fear of job loss due to industrial robots is unjustified as only fewer than 10% of jobs can be fully automated[7] with the remainder still being occupied by human workers. Optimistically, artificial and robotic intelligence is thought to create more jobs, not mass unemployment, as long as innovation is guided responsibly.[8] A study by the European Centre for Economic Research (ZEW)[9] argues that the two aspects, the drop in unemployment and the increase in robots, are closely linked and that robots are creating new jobs and not leaving masses of unemployed workers behind, as many people would predict. The study confirms developments in Eastern Europe and Romania, where robotization has allowed unemployment to diminish and wage increases to take place. In this context, the number of robots installed per 10,000 employees in Slovakia and Slovenia is higher than the global average of 74 robots per 10,000 employees, with more than 130 units. The Czech Republic has a density of 100 robots per 10,000 workers, while Hungary has 60, and Poland 30 units per 10,000 workers. Romania still has

2 http://www.epochtimes-romania.com.
3 World Robotics Report 2016, www.ttonline.ro.
4 http://reports.weforum.org/future-of-jobs-2016.
5 Georgescu 2018. 3–18.
6 Georgescu 2018. 2–15.
7 http://www.hotnews.ro.
8 https://www.wall-street.ro.
9 https://www.zew.de/en.

a low density of 15 robots per 10,000 employees and needs over 10,000 robots in the coming years to remain competitive in the region.[10]

At the World Economic Forum (Davos) meeting in January of 2018, the adaptability of companies to the new and revolutionary challenge of artificial intelligence was discussed. What has been made very clear is that the Fourth Industrial Revolution will eliminate millions of jobs.[11] There is no doubt that the world of collaborative robots is on the rise, so labour law will have to distinguish between non-human workers (industrial robots) and human workers. Regulations in the field will evolve, meaning that provisions will be needed which will determine, at a minimum, what the relationship between the two categories of workers will be according to the specificity of the activity as well as other aspects.

Social security, if jobs are reduced due to re-technologization and introduction of AI implementations, could be offset by state-owned companies or by introducing indemnities, permanent social benefits to maintain a decent living standard for humanity. Thus, man should no longer be concerned about subsistence needs – shelter, hygiene, food, etc. – but should be able to develop his creative part, educate and teach new generations in this regard, thus finding time for new inventions, for new solutions, new experiments and discoveries.[12]

It remains to be seen how collective bargaining will affect the future of industrial robots: can unions or employees' representatives force the employer to use only a limited number of industrial robots? Will employers be able to replace the work of human workers with industrial robots in case of strikes or the absence – for other objective reasons – of workers? These questions, as well as many others, are awaiting a firm response from the legislator, the only one able to ensure a reasonable balance.

The Fourth Industrial Revolution, or Industry 4.0, which 'blurs the boundaries between physical, digital and biological spheres',[13] starts with the already existing digital revolution, which will advance the economy in new, surprising directions based on robotics, artificial intelligence, nanotechnologies, biotechnology, the Internet of Things, 3D printing, or autonomous vehicles, and so industrial relations will change as robotization progresses.[14]

In this context, the European Parliament Resolution of 16 February 2017 containing recommendations to the Commission on civil law on robotics (2015/2103 (INL))[15] should be noted. According to this document, the implications are direct:

10 https://www.universal-robots.com/ro/.
11 Georgescu 2018. 3–16.
12 Georgescu 2018. 29.
13 Schwab 2016.
14 http://adevarul.ro/tech/stiinta/o-dezbatere-necesara-privind-viitorul…/index.htm.
15 http://www.europal.europa.eu.

— on jobs for, as the document notes: 'the widespread use of robotics may not automatically lead to the replacement of jobs, but less skilled jobs in intensive occupational sectors could be more vulnerable to the expansion of automation' and

— on the structure of society by excessive polarization and increasing the gap between the rich and the poor, as stated in the following way: 'in the face of growing divisions of society with a declining middle class, it is important to bear in mind that the development of robotics can lead to an acute concentration of wealth and influence in the hands of a minority'.

As far as Romania is concerned, we note a national interest in these topics, which is materialized by the adoption in 2015 of the National Strategy on the Digital Agenda 'Romania 2020',[16] which, although does not address the issue of robotics directly according to the European Parliament, has an important economic component through Action 3 – eCommerce, Research, Development, and Innovation in ICT. It is estimated that 'the implementation of measures under Action 3 will generate by 2020 an estimated impact on the Romanian economy of around 3% to GDP and 2% to jobs.'[17]

The importance of this Strategy is once again reinforced by the Governance Program 2017–2020, which has a distinct component called Communications Policies and Digital Convergence. 'Fast and unlimited access to information and facilities of the information, communication and computing tools for the better use of human energies, the modelling of a fair and creative society that contributes to the economic development and the increase of Romania's competitiveness'.[18]

Digitalization is also one of the pivotal concerns of the European Union. The Digital Single Market, an integral part of the 2020 Strategy, is built around new principles and ideas, such as 'digital inclusion'[19] (correlated with social inclusion), ideas designed to allow all categories of people to take part in the technological changes that digitalization brings with it.[20]

Europe is considering granting rights and responsibilities to robots with artificial intelligence. The European Parliament adopted a resolution[21] in 2017 providing for a special legal status of 'electronic people', that is to say, for autonomous robots. 'We are in the age of human intelligence along with the artificial one', argues the report. Such a new category of legal subjects that might have rights and obligations would be added to traditional ones, legal entities, and individuals who might be present at a certain moment in the labour market.

16 Government Decision No 245 of 7 April 2015, published in the Official Gazette of Romania No 340 of 19 May 2015.

17 https://www.antena.ro/…/a-patra-revolutie-industriala-este-posibila-o-robo-apocalipsa.

18 http://www.cdep.ro/pdfs/guv201706/Program_de_Guvernare.pdf.

19 https://ec.europa.eu/digital-single-market/en/digital-inclusion-better-eu-society.

20 Dimitriu 2016. 446.

21 http://www.europarl.europa.eu/news.

It was said[22] that 'the humanoid robot Sophia, the first robot who acquired citizenship (Saudi Arabia decided [in this way] in October 2017), is considered a thing, and not a person, and must be dismantled and brought into luggage to travel by aeroplane, for example, and granting human rights to humanoid robots, even if they are much reduced at an early stage, would be a major error in the thinking of any legislator. It will be just a step towards eliminating people…' This robot was recognized as benefitting of personhood by a fiction of the law.[23] If we take as a basis the idea that a humanoid robot is a man-made thing, however, it should in no case be regarded as a legal entity, not even based on a legal fiction, as was the case in the nineteenth century with legal entities, entities made up of human individuals.

Even if it can be argued that humanoid robots cannot function without the software of the physical (natural) person that creates them, there is still the fear that they will be able to 'update' to the point that they will no longer need software and thus recur to the elimination or dominance of human intelligence. Creating a register for intelligent autonomous robots, as proposed by the European Commission, would only solve the non-contractual liability issue in the case of damage caused by the intelligent robot, which would have to be corrected at the cost of the owner. It cannot be considered a document for the recognition of an 'electronic person' as a distinct subject of law.

However, the stronger presence of intelligent robots, like in the case of Germany, cannot be ignored. For example, that country boasted the largest number of robots per 10,000 workers in Europe, namely 309.[24] Employers would be tempted to use more and more of these non-human workers because they can work without being limited to a work schedule, are unable to make use of union claims, and need not benefit from health and safety measures at work.

If an artificial intelligence application achieves consciousness, 'we can say that it can do legal deeds and acts, manifesting its external consent in one way or another, by written or by mutual consent, depending on the nature of the legal deed or the act. Therefore, it may even conclude contracts, thus replacing even the manager of a company.'[25]

There were between 1.5 and 1.75 million industrial robots worldwide in 2017, according to the International Federation of Robotics.[26] The car industry employs about 39% of them, followed by the electronics industry (19%), the metal products sector (9%), and the plastics and chemicals industry (9%). Romania might well be the first country to have an artificial intelligence as ambassador, designed to

22 Dobozi–Colțan.2018.
23 Georgescu 2018. 3–18.
24 http://www.hotnews.ro.
25 Georgescu 2018. 3–20.
26 World Robotics 2017 – Service Robots; https://ifr.org/.

answer questions about Romania to foreigners and to make recommendations for visiting certain tourist areas in the country, talking about people's habits and their way of life.[27]

In areas exposed to industrial robots between 1990 and 2007, both employment rates and wages decreased significantly compared to other areas,[28] suggesting two solutions: 1. vocational reorientation programs for those whose jobs are taken over by robots and 2. reforming the education system. In a very short time, jobs may suffer. Though there will be no question of a redress of this situation in the future, a short-term (possibly drastic) reduction of jobs, due to the implementation of artificial intelligence in social life, is imminent, depending only on its ability to learn and adapt.[29]

The optimistic view that 'robots will have a complementary role and will not replace humans'[30] was in the past criticized by personalities such as Stephen Hawking and Elon Musk, who warned that artificial intelligence is a fundamental risk to the existence of human civilization.[31] Companies will, however, prefer artificial intelligence because there are much lower costs, and efficiency increases considerably due to its use. AI does not get tired, does not need a meal break, does not need rest, and does not have to work 8 hours a day; moreover, it does not need salary[32] burdened by taxes and social security costs.

Industrial Revolution 4.0 is a natural step in the evolution of humanity, a new challenge for human civilization, which should not restrain itself from using robots in economic activity. They will never be able to fully replace human intelligence; artificial intelligence, even if superior to the human intellect, will always be dependent on the latter, which will have the lead role. Replacing human labour with robots is no longer just a discussion, it is a reality, it is not just a science fiction issue. It is something society should think about and anticipate by updating legislation and social protection in some way or another in the interest of the people.

Some examples of the replacement of human workforce at an international level are already relevant: e.g. a New York Hotel, Yotel, which is fully automated and assisted by AI.[33] It has an automatic check-in and check-out, adjustable and comfortable, motorized bedding that folds to provide the client with extra room space, a robot permanently prepared to help customers with luggage, etc. China

27 Georgescu 2018. 3–20.
28 Acemoglu–Restrepo 2018.
29 Georgescu 2018. 3–20.
30 http://www.zf.ro/…/era-coboților.
31 Sisea 2017.
32 Georgescu 2018. 2–17.
33 https://www.youtube.com/watch?v=U81M7SjZjWY; https://www.tripadvisor.com/LocationPhotoDirectLink-g60763-d2079052-i75110632-YOTEL_New_York-New_York_City_New_York.html.

announced in November 2017 the planning of the opening of police stations without human staff, fully automated and assisted by AI.[34]

Another example of the AI that took the place of people is Amelia, who works at a UK local council. Amelia is scheduled for customer service and administration; she can analyse natural language, understand the context, apply logic, learn, solve problems, and even feel emotions.[35]

These applications, and others already in the research pipeline, are providing us with a preview of things to come. If labour law is unable to keep pace with technology or fails to consider the needs of human workers in the coming age, dystopian conditions may arise. A well-built legislative framework for robot–human interaction in the workplace may, on the other hand, herald a bright future.

References

ACEMOGLU, D.–RESTREPO, P. 2018. *Artificial Intelligence, Automation and Work*. https://www.nber.org/papers/w24196.

COLGATE, J. E.–PESHKIN, M. A.–WANNASUPHOPRASIT, W. 1996. Cobots: Robots for Collaboration with Human Operators. *Proceedings of the International Mechanical Engineering Congress and Exhibition, Atlanta, GA, DSC* 58: 433–439.

DIMITRIU, R. 2016. *Labour Law. Anxiety of the Present*. Bucharest.

DOBOZI, V.–COLȚAN, T. *Drepturi civile pentru roboții umanoizi?* https://www.hotnews.ro/stiri-specialisti_stoica_si_asociatii-22402441-drepturi-civile-pentru-robotii-umanoizi.htm.

GEORGESCU, L. 2018. What Is and How to Use Artificial Intelligence (I). *Romanian Journal of Labour Law* 3–20.

SCHWAB, K. 2016. *The Fourth Industrial Revolution.* Geneva.

SISEA, C. 2017. *Doi roboți au inventat un limbaj propriu şi au speriat Internetul. Este cazul să ne îngrijorăm?* http://www.ziare.com/internet-si-tehnologie/tehnologie/doi-roboti-au-inventat-un-limbaj-propriu-si-au-speriat-internetul-este-cazul-sa-ne-ingrijoram-1476258.

Online Sources

Robots: Legal Affairs Committee Calls for EU-Wide Rules. http:// www.europarl.europa.eu/news.

World Robotics 2017 – Service Robots. https://ifr.org/.

World Robotics Report 2016. www.ttonline.ro.

34 https://futurism.com/chinas-ai-police-station-humans/; https://thenextweb.com/artificial- intelligence/2017/11/09/china-is-building-a-police-station-powered-by-ai-nothumans/.

35 https://www.mirror.co.uk/news/uk-news/robot-amelia-who-can-sense-8215188.

http://adevarul.ro/tech/stiinta/o-dezbatere-necesara-privind-viitorul…/index.htm.
http://reports.weforum.org/future-of-jobs-2016.
http://www.cdep.ro/pdfs/guv201706/Program_de_Guvernare.pdf.
http://www.epochtimes-romania.com.
http://www.europal.europa.eu.
http://www.hotnews.ro.
http://www.zf.ro/…/era-coboților.
https://ec.europa.eu/digital-single-market/en/digital-inclusion-better-eu-society.
https://futurism.com/chinas-ai-police-station-humans/.
https://thenextweb.com/artificial-intelligence/2017/11/09/china-is-building-a-police-station-powered-by-ai-nothumans/.
https://www.antena.ro/…/a-patra-revolutie-industriala-este-posibila-o-robo-apocalipsa.
https://www.mirror.co.uk/news/uk-news/robot-amelia-who-can-sense-8215188.
https://www.tripadvisor.com/LocationPhotoDirectLink-g60763-d2079052-i75110632-YOTEL_New_York-New_York_City_New_York.html.
https://www.universal-robots.com/ro/.
https://www.wall-street.ro.
https://www.youtube.com/watch?v=U81M7SjZjWY.
https://www.zew.de/en.

Acta Univ. Sapientiae, Legal Studies, **8**, 2 (2019) 253–268

DOI: 10.47745/AUSLEG.2019.8.2.08

What Will Robot Laws Look Like? The Code of AI and Human Laws

Zsolt Ződi
Senior Research Fellow
Institute of Information Society
National University of Public Service, Budapest
E-mail: zodi.zsolt@tk.mta.hu

Abstract. The author aims to present in the course of this study the possible future interactions between laws and the behaviour of artificial intelligence. Firstly, the theory of code is presented as well as the debate regarding the aptitude of laws to represent a means for the control of machine behaviour either directly or, as is more likely, when embedded in code. Secondly, the author analyses the consequences of the emergence of 'robot law', the ways in which a mixed, two-, or possibly three-tiered normative system is arising. In such a system, human-readable law and robot law are likely to diverge and even possess different characteristics such as an added degree of instability in the case of robot law. The author analyses the difficulties posed by transitioning between these systems and those of endowing machines with behavioural concepts such as ethics and unbiased action, problems compounded by the inherent opaqueness of the processes which underpin artificial intelligence. Finally, the author raises the possibility that codes designed to regulate human–machine interrelationships in and of themselves may constitute the beginning of a new, supranational legal system, with the platforms employing such codes transformed into quasi-sovereign entities.

Keywords: artificial intelligence, normative regulation of behaviour, laws governing artificial intelligence, laws and machine learning, opaqueness of artificial intelligence decision making

1. Introduction

Since Asimov's three laws of robotics,[1] we have taken for granted that robots (and artificial intelligence) should *somehow* be regulated. It is also a commonplace that this regulation should look something like Asimov's laws, at least in one respect – namely, that they should impose a 'duty' on the robots (or on their developers?) *to do no harm* to human beings or, in other words, to comply

1 Asimov 1991. 37.

with the same laws people should follow. Most of the ethical guidelines that have been collected on a dedicated website[2] ultimately have this characteristic feature although some of them express it in a very detailed and sophisticated way.

The interrelationship between law and AI (robots),[3] however, is a lot more complicated than these ethical standards might suggest. First, it is quite obvious that robots are not governed by legal rules. They are controlled by algorithms, and algorithms are not expressed in (or rather embedded in) natural language as laws are; they are *codes*, collections of zeros and ones, often intertwined with some form of hardware. In other words, the 'transporting agent' of a code – unlike a law – is *not* the language itself. This raises serious questions about the extent to which humans can predict and understand them. A further and even more serious challenge is the *translation* of laws and ethical standards (values and other aspects) into codes and vice versa. And even if we properly translate laws to algorithms, we must still keep our natural language-based laws because we need them. This will surely lead to a double normative system. Within a short period of time, a further issue will also arise because, unlike laws, algorithms do not have to be fixed entities as robots are not confused by rules in constant flux the way people are: for them, stability and foreseeability are not indispensable, so it makes no sense to limit their capacity with rigid rules. We must let them adapt to changing circumstances – so long as they stick to certain high-level standards. Finally, the codes – a great part of them AI code – which are and will be running within the large-scale platforms and other services that are and will be central to our everyday life are opaque to us either because they are proprietary or because they are too complicated for a comprehensible explanation. Sometimes even their programmers do not see clearly how they function in individual cases – as in the case of neural networks – because they self-train themselves and constantly change. The opaqueness of the code limits human influence over them: and this is also problematic for nation-states. Although governments attempt to regulate platforms, and the platforms obey them, ultimately governments are not able to control the codes themselves.

In short: codes – the laws of robots and AI – show greater differences than similarities to our language-based laws. This paper attempts to present some of the problems which follow from this. The structure of the paper is as follows: in part I, I will recapitulate the theory of code – mainly following the arguments of the debate that has been going on since the middle of the 90s. In part I, I will also demonstrate how algorithms and laws differ. Part II deals with some of the consequences of the emergence of algorithms as a special means of behavioural control: their non-linguistic character, the consequences of the two parallel

2 AlgorithmWatch 2019.

3 In this Article, I will use 'robots', 'AI', and 'agents' synonymously.

normative systems (algorithms and laws), the problem of dynamic (self-training) rules versus predictability, the difficulty of coding value choices, and the global and secret nature of platforms' codes versus local, state-issued laws.

2. The Debate around Code as a Means of Behavioural Control

Soon after the emergence of the Internet ('cyberspace', as it was then called), Johnson and Post[4] raised the point that it 'requires a system of rules quite distinct from the laws that regulate physical, geographically defined territories'.[5] Already at the end of the 90s, Reidenberg[6] argued that in cyberspace rules are embedded into systems and technology.

> [F]or network environments and the Information Society, however, law and government regulation are not the only source of rulemaking. Technological capabilities and system design choices impose rules on participants. The creation and implementation of information policy are embedded in network designs and standards as well as in system configurations.

This became the leitmotif in Lawrence Lessig's seminal book, first published in 1999,[7] and also in the second edition of 2006.[8] Lessig differentiates four types of regulation – or four types of constraints to human behaviour: laws, norms, markets, and architecture. Laws are well-known to us. What Lessig calls 'norms' are mainly customs – widely accepted ways of (or beliefs about) behaviour. Markets regulate behaviour via supply, demand, and prices (the costs of resources). The fourth control is architecture: the design of the outer physical world.

> We can call each constraint a 'regulator', and we can think of each as a distinct modality of regulation. Each modality has a complex nature, and the interaction among these four is also hard to describe. (…) The code or software or architecture or protocols set these features, which are selected by code writers. They constrain some behavior by making other behavior possible or impossible. The code embeds certain values or makes certain values impossible. In this sense, it too is regulation, just as the architectures of real-space codes are regulations.[9]

4 Johnson–Post 1996.
5 Johnson–Post 1996. 1367.
6 Reidenberg 1997–1998. 554.
7 Lessig 1999.
8 Lessig 2006.
9 Lessig 2006. 124–125.

These regulators are quite different in many respects, but one of the most spectacular differences between them is their 'transporting agent'. Laws are encapsulated – i.e. they are formulated – in a sometimes difficult but still intelligible human language. Customs are typically embedded into human behaviour, but sometimes this widely accepted behaviour can also be underpinned by written rules, as is the case with diplomatic etiquette. The market as a regulator lies somewhere in between: supply and demand as expressed in the form of prices are mainly the result of an 'invisible hand', but they can be subject to rules expressed in a written form, for example, in the way 'fair commercial practices' are enforced and unfair practices are prevented by the competition authorities. Finally, architecture – like a fence, a wall, a speed bump, or an anti-theft tag – is part of the physical world. We can see here (paraphrasing Austin)[10] that while in law things are done with words, in the case of norms (customs) they are done with human actions, and in the case of codes we do things with 'things': we form the physical world in such a way that we shepherd humans in certain directions.

What makes this relevant to us, again, is that in cyberspace architecture is the code. It is not self-evident that code (the algorithm, or software) is a 'thing' or is part of the physical world. But amongst computer scientists it has long been a truism that:

> (H)ardware and software are logically equivalent. Any operation performed by software can also be built directly into the hardware, preferably after it is sufficiently well understood. As Karen Panetta Lentz put it: 'Hardware is just petrified software.' Of course, the reverse is also true: any instruction executed by the hardware can also be simulated in software.[11]

In short: the only way of controlling AI and robots is through the code. Robots cannot be controlled by laws, customs, or by market constraints. (Even trade robots, which seem to react to prices and other market variables, are ultimately controlled by codes and not directly by the prices on the stock exchange.)[12]

Controlling human behaviour with codes has many unique features. Code is not 'normative' in the classical sense, as neo-Kantian legal philosophy perceives normativity, because it is simultaneously a normative structure and therefore part of the intelligible world, and at the same time it has a physical manifestation. Code is a pre-fixed causal relationship. It is part of nature because it can be a number of switches and gates in the physical world, but it is also a self-training – or self-

10 Austin 1962.

11 Tannenbaum 2006. 8.

12 The phenomenon is regulated in the MiFID 2 Directive (Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU), which rules that high-frequency algorithmic trading systems should comply with the legal requirements of MiFID 2.

adaptive – algorithm (as are many AI-s), which can make surprising decisions and can produce end results that seem non-deterministic. It acts in an unforeseeable way (and also freely?). So, the difference between *is* and *ought to be* disappears in the case of code.

Since code has no linguistic manifestation – it is a virtual architecture –, its translation to human language is non-trivial. In most cases, our everyday narratives cannot be converted to code, and vice versa. Just try to explain in simple language how the Google ranking algorithm works (besides the simple fact that it puts the 'more relevant' items on top). How can the functioning of such a complicated code – written by hundreds of programmers over several decades and improved by billions of searches every day – be explained in human language?

3. Consequences of the Emergence of Robot Law

These characteristics of code have some serious impacts on the future concept and functioning of law, the future of legislation, the judicial interpretation of law, and the way ordinary citizens comply with legal rules. I will demonstrate these impacts on five fields.

3.1. A Mixed Normative System

The first and most spectacular result of the emergence of behavioural control through code is that, while in certain fields where machines entirely take over tasks performed by humans and law will disappear, in certain fields where machines and people 'live together', there will be a double – or even triple – legal system.

Machines are not limited to taking only 5–6 circumstances into consideration when they make decisions, as we humans are, and they can use, process, weigh up, and rely on hundreds or even thousands of parameters, each representing a particular circumstance in a real-life situation. Therefore, the temptation is huge to 'make laws more automation-friendly by specifying them differently and in more detail'.[13] In certain fields, lawmakers will be unable to resist this temptation, and they will create legal rules in such a way that they can be easily implemented through machine codes.

Another interesting consequence could be what McGinnick and Wasick have pointed out: the emergence of a new type of norm we might call 'dynamic rules'. 'Dynamic rules (…) set law's algorithm in silicon, permitting changes in law to occur only in response to previously specified information'.[14] Dynamic rules adapt to circumstances and change automatically in response to changes in external

13 Froomkin 2016. xix.
14 O'McGinnis–Wasick 2014. 997.

information. They 'change (…) by the application of prescribed formulas to new facts as those facts become available.'[15] The authors contend, following Kaplow's thesis,[16] that 'rules are generally more expensive to create, but then generally have lower enforcement costs'. Dynamic rules in this context are the 'standards' created for machines because they have the characteristic advantages of standards (flexibility and cost-effectiveness) without the drawbacks. Even though the authors draw radically different conclusions from the emergence of these new types of rules than I do (namely that a new 'supercharged' legal research system is needed which allows citizens to access legal information more directly), they are right that the phenomenon of self-adapting rules will become a reality in a few years. And to be even more utopian, just try to imagine what happens if these algorithms become self-teaching, self-developing ones, or if they can communicate with each other and learn from each other's experience.[17] The rules which are developed by machines in not easily foreseeable directions are a current reality in self-driving cars.

These self-trained, dynamic rules will be extremely complex and will very probably not be intelligible to humans anymore. We will then need a translation, a parallel system of rules. This might create a double or even triple legal system: one complicated and 'quantified' code for machines, containing thousands of variables and formulas, another, still rather complicated one for lawyers to handle complaints and to serve as a basis of judgement in the case of conflicts, and perhaps yet another system for the ordinary citizen. The simplified explanation of (over)complicated legal rules is not something esoteric: it already exists and is very successful. I am thinking here of the Creative Commons movement's pictograms.[18] Creative Commons is an initiative by Lawrence Lessig, and it primarily aims to promote free licences for copyrighted material, and therefore it offers model licence agreements. For ease of understanding, they have introduced symbols that represent the main rights (rules) within the licence agreement. For example, the crossed dollar sign means that the work can be freely used but not for commercial purposes. In short, these pictograms represent the most important rules within licence agreements. They are, in effect, the compressed and human-friendly versions of very technical licence agreements that can extend to a hundred pages. This method of representing complicated rules in a simple form will become quite ubiquitous in a world where most of the rules are written for machines.

At the same time, law will disappear altogether from some fully automatized fields. Consider the example of driverless cars. Great efforts have recently been made to teach driverless cars to recognize and interpret road signs originally designed for humans. But consider the 'Danger of fallen rocks on the road' sign.

15 O'McGinnis–Wasick 2014. 994.
16 Kaplow 1992.
17 Giarratana 2016.
18 Creative Commons 2019.

For humans, this means that they should be careful, probably slow down, and watch the road ahead closely. But what does it say to the driverless car's algorithm? Would it be scanning the road ahead more carefully? 'More carefully' makes no sense for a machine since it is continuously scanning the road anyway. For the software of the driverless car, any rule should be translated into a command *that changes the output*, i.e. reducing speed by not pushing the gas pedal, or using the break, or turning the wheels, etc. Just like the example mentioned here, from this perspective, most road signs make little or no sense to the software, whereas a rule such as 'reduce speed to 30 km/h in 2 seconds' does. If there are only driverless cars on the road, the traditional traffic rules will disappear. They will be superseded by direct electronic signals and messages. The new, more effective highway code for driverless cars will be a command system comprising signals and protocols which will directly determine the outputs of the driverless cars. Vague pictures which require human interpretation will stay but only as courtesy information for the passengers so that they can understand why the machine is doing certain things.

3.2. The Problem of Translation

The translation of legal rules to codes, and vice versa, will be a challenging issue. At first sight, codes are more transparent than laws because they are not corrupted by the fuzziness of everyday language, and they comply with the rules of logic. On the other hand, if there are too many logic gates, junctions, layers, and rules within a system, its functioning starts to become unpredictable and non-explainable in human language, that is – as Frank Pasquale terms it in his seminal book –, 'black-box like'.[19]

As long as architecture is visible, and codes are simple predictable restraints (like the rules of the PC game called Solitaire in Susskind's famous example),[20] codes will function as physical architecture. This already starts to become too complicated when algorithms start to manage longer processes or to make decisions on the basis of (sometimes rather difficult) decision trees. These semi-automatic and automatic decisions have already been part of legal ecosystems for decades, for example, in the form of traffic law enforcement systems.[21] Likewise, all around the world authorities are using expert systems and automatic document generation tools. Although such algorithms generate relatively easy decisions in simple problem situations (based on a few numeric parameters), they already raise certain questions.

19 Pasquale 2015.
20 Susskind 2008. 141.
21 Blackburn–Gilbert 1995.

In these cases, the problem is, as Bryant Walker Smith[22] points out, that the language spoken by lawyers and by technical staff is different. In a sense, this problem is rooted in the familiar issue of the ambiguity of legal language – which is part of everyday language – when compared to any meta-language of logic or maths. This was already highlighted by Lee Loevinger, founder of the Jurimetrics movement, in 1949:

> The difficulty is that we have no terms to put into the machines, as the scientists have numbers and symbols. Legal terms are almost all vague verbalizations which have only a ritualistic significance. (…) [T]he choice of legal terms to describe an act is certainly not a 'logical' operation. Where it is not purely arbitrary, it is, at most, intuitive.[23]

So, no matter how absurd it sounds, Liza Shay and her co-authors are right when they state that 'robots dream of electric laws'.[24] The authors performed an experiment where 52 programmers were assigned the task of automatic speed limit enforcement, and, even in this relatively easy case, where rules are narrow and straightforward, the number of mock tickets issued (i.e. legal consequences computed by the algorithms) varied to a very substantial extent. Clearly, the programmers interpreted the rules and the possible factual situations in very diverse ways at certain points. But simply coding traffic rules for AI is a challenge.[25] The same is true when the rules of GDPR need to be translated into a code on Facebook.[26]

The problem of translation is a serious challenge in the opposite direction too, when the decisions of the computer (AI) should be explained in some way. People need explanations for decisions, and the more a decision affects their lives, the more they demand a justification. This is the reason why the rules of automated decision making include the right to an explanation.[27] And explanation – even though some legal cultures prefer to refer to it as a logical subsumption – is more like storytelling. Does this mean that we have to teach our machines storytelling? I will return to this in more detail in the next point.

22 Walker Smith 2016. 78–101.
23 Loevinger 1949. 471–472.
24 Shay et al. 2016. 274.
25 Carp 2018, Prakken 2017.
26 Houser–Voss 2018.
27 Articles 13(2)(f), 14(2)(g), and 15(1)(h) of the GDPR require data controllers to provide data subjects with information about 'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'. At the time of the birth of GDPR, there was a lively debate in the literature on whether these texts are rules of 'the right to explanation' or not. See: Selbst–Powles 2017.

This phenomenon really becomes ponderous when more complex – legal rule-based – decisions are supposed to be made by artificial intelligences. This is the case, for example, when decisions made by a robo-advisor need to be explained under MiFID 2,[28] or there is a requirement to explain the decisions of the AI which controls content on Facebook.[29] Compared to 2006, when the 2nd edition of Lessig's book came out, the world of codes has expanded enormously because of the large-scale platforms such as social media, online shops, search engines, and matching software like Uber and Airbnb. These platforms have started to dominate our online activities, and they are all based on codes. On platforms, where there is no 'real' physical architecture but only a virtual architecture, the limits and boundaries of (virtual) space, the ways of speaking and acting (online), the channelling of attention, the display of the outer world, and 'life and death' decisions (such as allowing or banning an account) are all determined by codes. A platform's main architecture is its code, and codes have started to control spheres (intimate space, private life, and the means of cognition), what they never did before.

3.3. Values in the Code

A further consequence is that, for these algorithms to work, we have to make more explicit the choices underlying our rules and assign values to them in more explicit ways. The trouble is that value choices are sometimes very hard to justify or even to express.

Modern legal systems are built up in a systemic way. Detailed rules are first based on codes and finally on values and principles, codified in the codes or in the constitution. Although the common law systems' architecture is slightly different, this pyramid of abstraction exists there, too. One can think of constitutional principles of the freedom of speech or special principles of criminal justice such as 'nullum crimen sine lege', and so on. Administrative decision making and judicial interpretation rely on these principles and policies very heavily.[30]

One might argue that codes and algorithms, as well as algorithmic decision making, have nothing to do with these principles. Still, it seems that in the last few years the 'morality of the codes' has become one of the most important debates within 'robot law'. There is a website which collects the ethical codes of AI (algorithms) from all around the world, and there are more than 80 of these ethical codes available. In most of them, we see recurring requirements: codes should be transparent, codes must not discriminate, decisions made by the codes should be explainable, etc.[31]

28 ESMA-EBA-EIOPA Report 2018. 9.
29 Bickert–Fishman 2017, Macdonald et al. 2019.
30 Dworkin 1967.
31 AlgorithmWatch 2019.

The risks associated with making high-level values or principles explicit are even more visible when we want to quantify the principles of justice – the ultimate value in law. There are several ways to organize a judicial system, but all of them reveal that justice is Janus-faced: it requires each case to be decided on its own merits, on the basis of the special, individual, and unrepeatable circumstances inherent to the event, but at the same time it also requires similar cases to be treated alike. If a decision-making algorithm can take a practically endless number of parameters into consideration, there is a strong temptation to use all of these parameters (as justice must account for all the relevant considerations). But then, after a while, there will be no more 'similar' cases. Each case will be judged on its own merits, which ultimately – I imagine – will undermine our sense of justice.

One can argue that even this inner tension can be quantified with the help of, say, vector maths or cluster analysis. In the former case, 'vectors' are the circumstances that have to be taken into account, while 'clusters' are the groups of cases that have to be treated 'equally'. I am unable to assess at this stage the value or feasibility of these methods. Both might work. But whatever the result is, it is still true that in these cases value choices should be made explicit and should be somehow quantified. With human decisions, we sometimes accept strange decisions, especially if the decision maker has great authority and provides a valid reasoning. I have doubts as to whether machines can provide acceptable reasoning. People are rationalizing rather than rational creatures. Persuasive reasoning is more important than the rational decision itself. The absurdity of algorithmizing high-level value choices becomes apparent when we go through the 'moral machine' test.

A further manifestation of the representation of values, principles, and policies in code is the question of flexibility, or equity, or mercy. Karnow formulated this question in terms of discretion: 'How much human discretion should be built into an automated law-enforcement system?'[32] Another expert in the field, Elizabeth Joh, when tackling the discretion problem (and recognizing that human policemen do not enforce every minor detail of the law), asks the following question: 'Would we live in a better world if police patrol robots enforced minor offenses much more frequently than human officers would in neighbourhoods accustomed to aggressive policing because they were directed to do so by their own artificial intelligence?'[33]

I think these puzzles confront us with two theoretical challenges. The first is the question of the extent to which law is an algorithm. This is crucial because robots can apply legal rules only when they are translated into algorithms (and this will

32 Karnow 2016. 51.
33 Joh 2016. 540.

be the case ever more frequently in the future). The second problem leads us to serious constitutional problems: the access to law and legitimacy issues.

3.4. Non-Transparent Functioning and Non-Explainable Results

Both in mixed and solely algorithmically managed ecosystems, the problem of understanding and explaining machine-made decisions will be a challenge. One might think that, in a certain respect, codes are more transparent because they are in line with the rules of logic, so they are not encumbered with the fuzziness and vagueness of ordinary language. But, on the other hand, their functioning can be so difficult – or, as the literature says, 'black-box-like' – that the end result of their operation cannot be explained in narrative-centred human language.

This issue had a sensational impact when software used by the courts sent a Wisconsin man, Eric L. Loomis, to prison because – according to the judgement of the algorithm – he showed 'a high risk of violence, [a] high risk of recidivism, [and a] high pretrial risk'.[34] Loomis obviously had no chance to study the algorithm and argue against it.

We have to recognize that all of these codes are based on an anthropomorphism. AI cannot understand these standards although it can understand and execute codes. But as soon as we start to operationalize these values, we encounter contradictions that cannot be represented on a code level. This is true in the case of a standalone principle (like the prohibition of discrimination[35] or justice itself), but it is even more spectacular when two or more of these principles are in conflict, which occurs quite frequently in constitutional law. This dilemma is clearly demonstrated by the 'moral machine' of the MIT, which simulates a moral dilemma in an imaginary situation where an autonomous vehicle must make a decision about causing harm.[36] A further problem is that in most cases AI systems are based on self-training algorithms, where the code is developing itself, and/or on 'big' data sets, where the data is produced in a spontaneous, uncontrolled way. In these cases, even the programmers of the code cannot foresee the output of the system, let alone explain the results.[37]

34 This was the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) software, which has been served since then as a deterrent example of machine bias. For the full story, see: Liptak 2017; for the machine bias in the case of COMPAS, see: Angwin et al. 2016.

35 What makes algorithmic bias a hopelessly complicated problem is 'indirect' discrimination, 'where an apparently neutral provision, criterion or practice would put persons of a racial or ethnic origin at a particular disadvantage compared with other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate' – Council Directive 2000/43/EC.

36 http://moralmachine.mit.edu/.

37 For the definition of AI, see: A definition of Artificial Intelligence 2019. A recurring element within the document is 'machine learning', which – at least at present – seems to be the key component of AI.

The real-life application of rules – the task of translating legal rules to machine commands – will be a task for the programmers, or rather it well might be that a separate profession will emerge, of 'legal knowledge engineers', as Susskind predicts.[38] They will 'organise the large quantities of complex legal content', analyse and 'distil' legal processes, and then embody these into computers systems.

Once again, one might say that there is no real problem here. Of course, programmers should interpret and translate human rules into the language of robotics. But that is already happening: legal rules are being implemented into, for example, ERP[39] systems, document generation software, tax software, and speed limit enforcement software. Apart from the interpretational problem I have already indicated, there is another issue: if we reach the period of law making for robots and dynamic rules, this profession will not only be the translator of human legal rules to algorithms but also the lawmakers and maybe the 'back-translators' of robot laws to human language. Or perhaps the back-translator will constitute a separate profession – who knows?

3.5. Platforms as Nation-States? Codes as New Legal Systems?

Another challenge is the tension between the international (non-national) character of codes and the national character of legal rules. This is not true for all codes but particularly true for those that affect our everyday life in the most profound way: the codes of platforms. Platforms organize our lives, we socialize on them, buy and sell on them, order different services on them, and so on. The codes of these platforms are international, while laws are artefacts of nation-states. This leads to a great deal of serious tension.

Firstly, it leads to a continuous battle between nation-states and platforms. Competition and data protection authorities levy more and more heavy fines on Facebook and Google.[40]

However, the even more frightening aspect of this phenomena is that given that the big platforms have their own (albeit virtual) borders, 'inhabitants', and power over their inhabitants: they have the main characteristics of a nation-state. As Julie Cohen puts it, platforms are 'emergent transnational sovereigns'.[41]

38 Susskind 2008. 272.

39 Enterprise Resource Planning software: integrated software tools that support enterprises in organizing their workflows. Many legal rules are embedded in these systems: labour law regulations in the HR module or tax regulations in the accounting module.

40 Just a few illustrations from the latest news: the Federal Trade Commission in the USA fined Facebook 5 billion dollars in 2019 (Glazer et al. 2019); the Italian competition authority in 2018 levied an 8.9 million pound fine (Hern 2018), and during the writing of this paper the Turkish Data Protection Authority fined Facebook 280,000 USD for a data breach (Turkey fines Facebook 2019).

41 Cohen 2017. 199.

There are many indications of this phenomenon: first, platforms build up their own, autonomous regulatory world. They have their own house rules,[42] which have so far been partly, at least in the case of Facebook, a secret document.[43] These rules comprise the definition of basic legal and constitutional concepts such as terrorism or defamation; and although there is a fierce debate as to whether Facebook is biased or not,[44] these rules visibly represent a liberal 'West Coast' set of values.[45] Platforms act like sovereigns, negotiate with governments and with competition and consumer protection authorities, and sometimes explicitly express the idea that 'in an age of nationalism' they want to be a 'trusted and neutral digital Switzerland', as Microsoft President and Chief Legal Officer Brad Smith declared in a conference.[46]

The last and most spectacular development of platforms on their move towards becoming sovereigns, and the transformation of their internal codes into an alternative legal system, is the introduction of the new cryptocurrency, Libra, by Facebook.[47] This cryptocurrency is again a global code because it runs on a blockchain. According to the official statement of the company, it is for those who have Internet access but no bank account because of the lack of financial infrastructure.[48] This global code may one day colonize another sphere involving millions of people.

The codes on these platforms are very often the trade secrets of the owners, as is the data they collect.[49] We cannot yet state that platforms are new nation-states, but their internal code, which decides what we can see from the outside world and how and what can we say, is a code-based competitive normative system. And the main goal of these platforms – no matter what they say – is still the generation of profit by 'datafying' and monetizing our personal data with the help of the code.

However, it is not only nation-states that are threatened by these platforms; there are other legitimacy or constitutional challenges raised by these codes. If there are two or three separate bodies of rules, then the millions of exact rules and dependencies and the precise maths of vectors and clusters will not be transparent for the ordinary citizen, and problems similar to that experienced by the man in Wisconsin may become quite common. What will the unity of law and the uniformity of courts' decisions look like in this new world?

42 https://www.facebook.com/communitystandards/.
43 See Hopkins 2017.
44 See, e.g., Senator Jon Kyl's report.
45 See, e.g., Conger–Frenkel 2018.
46 Conger 2017.
47 See Zuckerberg 2019.
48 Coming in 2020.
49 Pasquale 2015. 82.

References

ANGWIN, J.–LARSON, J.–MATTU, S.–KIRCHNER, L. 2016. Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks. *ProPublica*: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

ASIMOV, I. 1991. *I, Robot*. New York.

AUSTIN, J. L. 1962. *How to Do Things with Words*. London.

BICKERT, M.–FISHMAN, B. 2017. Hard Questions: How We Counter Terrorism. *Facebook News*: https:// newsroom.fb.com/news/2017/06/how-we-counter-terrorism/.

BLACKBURN, R. R.–Gilbert, D. T. 1995. *Photographic Enforcement of Traffic Laws. Synthesis of Highway Practice*. Washington D.C.

CALO, R.–FROOMKIN, A. M. (eds). 2016. *Robot Law*. Northampton (Massachusetts, USA).

CARP, J. A. 2018. Autonomous Vehicles: Problems and Principles for Future Regulation. *University of Pennsylvania Journal of Law & Public Affairs* 4: 81–149.

COHEN, J. E. 2017. Law for the Platform Economy. *U. C. Davis Law Review* 51: 133–204.

CONGER, K. 2017. Microsoft Calls for Establishment of a Digital Geneva Convention. *Techcrunch*: https://techcrunch.com/2017/02/14/microsoft-calls-for-establishment-of-a-digital-geneva-convention.

CONGER, K.–FRENKEL, F. 2018. Dozens at Facebook Unite to Challenge Its 'Intolerant' Liberal Culture. *The New York Times*: https://www.nytimes.com/2018/08/28/technology/inside-facebook-employees-political-bias.html.

DWORKIN, R. 1967. The Model of Rules. *University of Chicago Law Review* 35: 14–46.

FROOMKIN, A. M. 2016. Introduction. In: *Robot Law*. Northampton (Massachusetts, USA).

GIARRATANA, C. 2016. How AI Is Driving the Future of Autonomous Cars. *Readwrite.com*: 20.12.2016. https://readwrite.com/2016/12/20/ai-driving-future-autonomous-cars-tl4/.

GLAZER, E.–TRACY, R.–HORWITZ, J. 2019. FTC Approves Roughly $5 Billion Facebook Settlement. *The Wall Street Journal*: https://www.wsj.com/articles/ftc-approves-roughly-5-billion-facebook-settlement-11562960538?mod=e2tw.

HERN, A. 2018. Italian Regulator Fines Facebook £8.9m for Misleading Users. *The Guardian*: https://www.theguardian.com/technology/2018/dec/07/italian-regulator-fines-facebook-89m-for-misleading-users.

HOPKINS, N. 2017. Facebook Moderators: A Quick Guide to Their Job and Its Challenges. *The Guardian*: https://www.theguardian.com/news/2017/may/21/facebook-moderators-quick-guide-job-challenges.

HOUSER, K. A.–VOSS, W. G. 2018. GDPR: The End of Google and Facebook or a New Paradigm in Data Privacy. *Richmond Journal of Law & Technology* 25: 1–109.

JOH, E. 2016. Policing Police Robots. *UCLA Law Review Discourse* 64: 516–543.

JOHNSON, D. R.–POST, D. 1996. Law and Borders: The Rise of Law in Cyberspace. *Stanford Law Review* 48: 1367–1402.

KAPLOW, L. 1992. Rules versus Standards: An Economic Analysis. *Duke Law Journal* 42: 557–567.

KARNOW, C. E. A. 2016. The Application of Traditional Tort Theory to Embodied Machine Intelligence. In: *Robot Law*. Northampton (Massachusetts, USA).

LAWRENCE, L. 2006. *Code 2.0*. New York.

LEENES, R.–LUCIVERO, F. 2014. Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design. *Law, Innovation and Technology* 6: 193–220.

LESSIG, L. 1999. *Code and Other Laws of Cyberspace*. New York.

LIPTAK, A. 2017. Sent to Prison by a Software Program's Secret Algorithms. *The New York Times*: https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html?_r=0.

LOEVINGER, L. 1949. Jurimetrics; The Next Step Forward. *Minnesota Law Review* 33: 455–493.

MACDONALD, S.–CORREIA, S. G.–WATKIN, A. L. 2019. Regulating Terrorist Content on Social Media: Automation and the Rule of Law. *International Journal of Law in Context* 15: 183–197.

O'MCGINNIS, J.–WASICK, S. 2014. Law's Algorithm. *Florida Law Review* 66: 1991–1050.

PASQUALE, F. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge (Massachusetts, USA)–London.

PRAKKEN, H. 2017. On the Problem of Making Autonomous Vehicles Conform to Traffic Law. *Artificial Intelligence and Law* 3: 341–363.

REIDENBERG, J. R. 1998. Lex Informatica: The Formulation of Information Policy Rules through Technology. *Texas Law Review* 76: 553–594.

RICHARD, N. M.–SMART, W. D. 2016. How Should the Law Think about Robots? In: *Robot Law*. Northampton (Massachusetts, USA).

SELBST, A. D.–POWLES, J. 2017. Meaningful Information and the Right to Explanation. *International Data Privacy Law* 7: 233–253.

SHAY, L. A.–HARTZOG, W.–NELSON, J.–CONTI, G. 2016. Do Robots Dream of Electric Laws? An Experiment in the Law as Algorithm. In: *Robot Law*. Northampton (Massachusetts, USA).

SUSSKIND, R. 2008. *The End of Lawyers?* Oxford.
TANNENBAUM, A. S. 2006. *Structured Computer Organization*. New Jersey.
WALKER-SMITH, B. 2016. Lawyers and Engineers Should Speak the Same Robot Language. In: *Robot Law*. Northampton (Massachusetts, USA).
ZUCKERBERG, M. 2019. https://www.facebook.com/zuck/posts/ 10107693323579671.

Web Sources

ALGORITHMWATCH. *AI Ethics Guidelines Global Inventory 2019*. https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/.
COMING in 2020 Calibra, A New Digital Wallet for a New Digital Currency. https://newsroom.fb.com/news/2019/06/coming-in-2020-calibra/.
COUNCIL DIRECTIVE 2000/43/EC of 19 July 2000 Implementing the Principle of Equal Treatment between Persons Irrespective of Racial or Ethnic Origin.
CREATIVE COMMONS Linceses and Examples. https://creativecommons.org/share-your-work/licensing-types-examples/licensing-examples/.

ESMA-EBA-EIOPA. *Report on Automation in Financial Advice*. https://esas-joint-committee.europa.eu/Publications/Reports/EBA%20BS%202016%20422%20(JC%20SC%20CPFI%20Final%20Report%20on%20automated%20advice%20tools).pdf.
HIGH-LEVEL Expert Group on Artificial Intelligence. *A Definition of Artificial Intelligence: Main Capabilities and Scientific Disciplines.* https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines.
SENATOR Jon Kyl's Report. https://fbnewsroomus.files.wordpress.com/2019/08/covington-interim-report-1.pdf?mod=article_inline https://www.theverge.com/interface/2019/8/21/20825899/facebook-kyl-audit-conservatives-clear-history; https://www.nytimes.com/2016/05/19/opinion/the-real-bias-built-in-at-facebook.html?smid=fb-nytimes&smtyp=cur&fbclid=IwAR3CbmLMuLDP7Vvb82e0RpcOHv_3Jk9JJzOwj0IvjY6jWdjIabwDysUm9Zo.
TURKEY Fines Facebook over Data Breach. *Hürriet Daily News*. http://www.hurriyetdailynews.com/turkey-fines-facebook-over-data-breach-147105, https://libra.org/home-page-2/.

Acta Universitatis Sapientiae

The scientific journal of Sapientia Hungarian University of Transylvania publishes original papers and deep surveys in several areas of sciences written in English.
Information about the appropriate series can be found at the Internet address
http://www.acta.sapientia.ro.

Acta Universitatis Sapientiae
Legal Studies

Sapientia University

Scientia Publishing House

ISSN 2285-6293
http://www.acta.sapientia.ro

Instructions for authors

Acta Universitatis Sapientiae, Legal Studies publishes studies, research notes and commentaries, book and conference reviews in the field of legal sciences.

Acta Universitatis Sapientiae, Legal Studies is a peer reviewed journal. All submitted manuscripts are reviewed by two anonymous referees. Contributors are expected to submit original manuscripts which reflect the results of their personal scientific work. Manuscripts sent to the journal should not be previously published in other journals and should not be considered for publication by other journals.

Papers are to be submitted in English, French, German, Romanian or Hungarian, in A4 format, electronically (in .doc or .docx format) to **the e-mail address of the executive editor: tnotari@kv.sapientia.ro**

Manuscripts should conform to the following guidelines:

The length of the papers should not exceed 7,000 words (respectively 3,000 in the case of commentaries and reviews) and manuscripts should be accompanied by a 200–250-word abstract with 3-4 keywords and with authors' affiliation. Tables and graphs, if any, should be prepared in black and white, should be titled, numbered and integrated in the main text. The list of references should appear at the end of the manuscripts.

Acta Universitatis Sapientiae, Legal Studies is published twice a year: in May and December.

For information on the citation guidelines see: http://www.acta.sapientia.ro/acta-legal/legal-main.htm

Contact address and subscription:
Acta Universitatis Sapientiae, Legal Studies
Sapientia Hungarian University of Transylvania
RO 400112 Cluj-Napoca, Romania
Str. Matei Corvin nr. 4.
E-mail: acta-legal@acta.sapientia.ro

Printed by F&F INTERNATIONAL
Director: Enikő Ambrus