ARTIFICIAL INTELLIGENCE IN THE ECONOMIC SECTOR

PREVENTION AND RESPONSIBILITY

Editors

MARIA JOÃO ANTUNES · SUSANA AIRES DE SOUSA

This book was developed in the scope of the Exploratory Project "Artificial Intelligence and Corporate Crime", integrated in the Research Area "Risk-Transparency-Litigation" and as part of the project "Societal Challenges, Uncertainty and Law: Plurality | Vulnerability | Undecidability" of the University of Coimbra Institute for Legal Research (UIDB/04643/2020), which is financed by national funds from FCT.

Edition
Instituto Jurídico Faculdade de Direito da Universidade de Coimbra | University of Coimbra Institute for Legal Research

Graphic design
Tipografia Lousanense, Lda.

Contacts
geral@ij.uc.pt
www.uc.pt/fduc/ij
Colégio da Trindade | 3000-018 Coimbra

isbn
978-989-9075-19-1

legal deposit
493703/22

doi
https://doi.org/10.47907/livro2021_4

Instituto Jurídico | Faculdade de Direito | Universidade de Coimbra

ARTIFICIAL INTELLIGENCE IN THE ECONOMIC SECTOR

PREVENTION AND RESPONSIBILITY

Editors

MARIA JOÃO ANTUNES · SUSANA AIRES DE SOUSA

1 2 9 0

INSTITVTO IVRIDICO
FACULDADE DE DIREITO
UNIVERSIDADE D
COIMBRA

Contents

Introduction – AI in the economic sector: prevention and responsibility

(https://doi.org/10.47907/livro2021_4c1)

Susana Aires de Sousa[1]

> 'What sort of things do *you* remember best?' Alice ventured to ask.
> 'Oh, things that happened the week after next,' the Queen replied in a careless tone. 'For instance, now,' she went on, sticking a large piece of plaster on her finger as she spoke, 'there's the King's Messenger. He's in prison now, being punished: and the trial doesn't even begin till next Wednesday: and of course the crime comes last of all.'
> 'Suppose he never commits the crime?' said Alice.
> 'That would be all the better, wouldn't it?'
>
> Lewis Carroll, *Through the Looking Glass*

I. A digital transition is happening in the economic sector[2]. New technology – machine learning, language processing, robotics, electronic platforms, blockchain, cognitive computing, quantum computing… –, although all these are at different stages of development, is already integrating innumerable economic and financial activities. However, technologically accelerated evolution, in parallel with the digitalization of markets and the massive creation of data, have together favoured the emergence of algorithms capable of extracting and structuring, from

---

[1] Assistant Professor of Law. Univ. Coimbra, University of Coimbra Institute for Legal Research, Fac. Law.

[2] Theo Lynn / John G. Mooney / Pierangelo Rosati / Mark Cummins (Editors), *Disrupting Finance. FinTech and Strategy in the 21st Century,* Macmillan, 2019; also, dedicated to the challenges of Blockchain and AI, the *Journal of Corporation* Law, Volume *46, Number 4 (2021*), https://jcl.law.uiowa.edu/volume-46-number-4

big data, relevant (and economically valuable) information[3]. Typical advantages of complex computerized systems, such as the enormous capacity for data analysis (already impossible for human intelligence) are now upgraded with new AI techniques, with predictive and prescriptive skills. This predictive ability makes AI algorithms particularly suitable for and efficient at performing several tasks such as compliance obligations, fraud detection, cyberattacks prevention or as a simple commercial tool (in customer service and assistance, for example).

These advantages have long attracted the attention of several stakeholders in the economic sector. The impact of AI on the financial and banking system is undeniable. As the financial sector navigates risky choices based on probability judgements, the most favourable scenario is fashioned for algorithms "to do their thing", e.g., credit risk assessment, market risk analysis, economic operations performance (calculating measuring and identifying risks, probabilities and strategies), and fraud detection, etc.

Exposed in recent decades, in the wake of the 2008 financial crisis, to enormous pressure from regulators in the pursuit of their business, banks have found solutions in the advantages offered by new technologies. "Banking has always been particularly open to technical innovation and progress"[4]. However, the appearance of AI has brought about a real revolution in the financial field. Just consider the veritable digitalization of the institution, with the emergence of fully digital banks, without any physical existence. The securities market has also undergone profound changes with the introduction of trading algorithms capable of automating and accelerating transactions, increasing efficiency, velocity and liquidity[5]. The threat of a systemic risk linked to the behaviour of algorithms is however a recurrent concern and should not be ignored.

---

[3] Anabela Miranda Rodrigues / Susana Aires de Sousa, «Algoritmos em contexto empresarial: vantagens e desafios à luz do direito penal», *Julgar 45*, p. 193--215.

[4] Pedro Maia, Part I, Chapter 1.

[5] Douglas W. Arner / Janos Barberis / Ross P. Buckley, «The Evolution of FinTech: A New Post-Crisis Paradigm?», University of Hong Kong Faculty of Law Research Paper No. 2015/047, UNSW Law Research Paper No. 2016-62, SSRN: https://ssrn.com/abstract=. Also Gregory Scopino, *Algo bots and the Law*, Cambridge University Press, 2020, p. 177

Technologies do allow costs to be cut, switching from people to algorithms, and do enable better risk management[6]. A risk analysis of operations and operators, in compliance with the obligations imposed by regulators, seems to offer means of detection or even prevention of financial fraud, "evaluating the best ways to protect their systems, their data, and ultimately their clients"[7]. It should therefore come as no surprise that algorithms, with their ability to analyse patterns and detect suspicious movements, have established themselves as a powerful tool for compliance and fraud prevention and detection. AI solutions, although expensive[8], promise automated continuous monitoring, relieving the company of the costs associated with self-regulation and, on the other hand, making it easier for the regulator to quickly access information in case of non-compliance[9]. Fraud can be a financial sign or transfer, whose irregularity – undetectable to the human eye – is easily identifiable or flagged by an algorithm capable of comparing and analysing big data.

FinTech (Financial Technology), RegTech (Regulatory Technology) and SupTech (Supervisory Technology) represent this digital shift, both on a practical as a narrative level: whether in the banking sector or in the field of capital markets, architecture, structure, management and operations have been profoundly altered by networks of computerized systems that guide countless digital movements and transactions[10]. In a subsequent step, artificial intelligence techniques have revealed

---

[6] Butler / Brooks, «On the role of ontology-based RegTech for managing risk and compliance reporting in the age of regulation», *Journal of Risk Management in Financial Institutions*, Vol. 11 (2018), p. 19 e ss.

[7] Saqib Aziz / Michael Dowling, «Machine Learning and AI for Risk Management», *Disrupting Finance*, Palgrave Macmillan, 2019, p. 43.

[8] Aldridge and Krawciw note that in recent years investment in financial technology has grown by 201% worldwide, cf. «Real-time risk: What investors should know about FinTech, high-frequency trading, and flash crashes» Hoboken, NJ: Wiley, *apud* E. Monaco, «What Fintech can learn from high-frequency trading)», *Disrupting Finance*, Palgrave Macmillan, 2019, p. 52.

[9] Cf. Tom Butler / Leona O'Brien, «Artificial intelligence for regulatory compliance: Are we there yet?», *Journal of Financial Compliance*, Vol. 3, N 1, 2019, p. 45.

[10] Douglas W. Arner / Janos Barberis / Ross P. Buckley, «The Evolution of FinTech: A New Post-Crisis Paradigm?», University of Hong Kong Faculty of Law Research Paper No. 2015/047, UNSW Law Research Paper No. 2016-62, SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2676553

themselves to be an auspicious instrument for monitoring transactions and consequently as a powerful tool in the investigation of fraudulent practices in the financial market. Financial cybersecurity has been in fact one of the sectors leading this field.

These aspects, among others, are covered in depth in the first part of this book, which is entitled *Prevention*.

Pedro Maia in "Intelligent Compliance" describes the compliance obligations felt by the banking sector and the way in which new technologies have allowed institutions to respond to these demands. The complexity of the compliance system to which banking institutions are subjected, as a part of a "legislative tsunami" unleashed by the 2007-2008 financial crisis are described and analysed in detail. Technology became a powerful instrument of compliance responding to a duty of risk identification and mitigation. At the same time, the author does not omit to warn us of the possible consequences and costs of unlimited trust being placed on algorithms: the exposure of the financial system to new and significant risks, making the system – again – more fragile.

Alexandre Soveral Martins, in the chapter "Algo-trading", unveils a set of reflections on algorithmic trading, pointing out both its advantages and volatility risks. The reaction to the risk of instability, leveraged by High Frequency Trading (HFT), forced regulators to act in order to ensure or determine the conditions of trust that are essential to the functioning of this market. Risky behaviours facilitated by HFT are also listed by the author. Many of these behaviours are associated with market manipulation such as ping orders, phishing, quote stuffing, spoofing, wash trading, slow traders, etc.

José Ricardo Marcondes Ramos, in the chapter "The use of Big Data and Artificial Intelligence to prevent and detect fraud", explores the ideas of digital forensics through the use of AI. This paper focusses on the discussion about the role that AI is already playing in fraud detection. The enormous amount of data collected and extracted from all sort of technologies and devices (computers, platforms, phones, smart watches, etc.) is feeding the development of this new type of digital forensics based on AI techniques. Big data allows supervised and unsupervised training (alongside other methods as social network

analysis) to transform AI into a powerful tool capable of identifying patterns of fraud or suspicious activities connected with financial fraud, money laundering, financing of terrorism, market manipulation or corporate crimes.

II. However, the use of AI comes with risks and costs, in particular risks connected with algorithmic unpredictability that may cause harm to protected interests, whether individually or collective owned (altered prices, market manipulation, manipulated advertising, privacy attacks).

Scholars have pointed out numerous examples of automated decision systems going wrong such as traffic accidents with automated driving systems[11], spoofing orders on the market securities, phishing threats favoured by the internet of things[12], or even racist or biased outcomes[13].

The risks signalled of AI may indeed materialize in harmful wrongdoing, bringing about the question of who is responsible for them. A new set of problems emerge. The dystopian nature of complex computational systems make imputational categories seem inadequate. From the perspective of the human or corporate person involved in the manufacture, programming or use of the system, the intervention of the machine renders, sometimes, the harmful event unpredictable. Are our responsibility systems and the existing models of liability adequate to respond to harmful events connected with algorithmic decisions?

---

[11] Mihalis Diamantis, Part II, Chapter 2.

[12] Steven Furnell, «Technology Use, Abuse, and Public Perceptions of Cybercrime», *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, Palgrave Macmillan (ed. Thomas J. Holt / Adam M. Bossler), 2020, p. 45 e ss.

[13] Alan Rubel / Clinton Castro /Adam Pham, *Algorithms and Autonomy. The ethics of automated decision systems*, Cambridge University Press, 2021, p. 137 e ss., point out some examples: in 2018, an Uber-automated driving system failed to recognize a bicyclist, whom it struck and killed; in 2012, the Target Corporation received international attention when, based on predictive analytics and an automated advertising system, it sent fliers targeting women seeking prenatal products to a minor before she had revealed her pregnancy to one of her parents; in 2017, the news organization ProPublica was able to use Facebook's automated system to make an ad buy targeting users with anti-Semitic affiliations.

One could argue that those risks may be diminished by creating more accurate machines with exceptional predictive capacities. As Aziz and Dowling emphasise, as the organisation and analysis of data becomes more targeted and focused through AI we are "able to accurately know in advance the risks, be they company, market, operational or credit risks"[14].

Accuracy demands data, big data. Constant monitoring (of agents, transactions, values, connections, financial movements, website visits) becomes one of the main sources of data collecting. This "surveillance" happens both in a limited environment (*e.g.*, the surveillance of employees[15]) or on a wide scale (*e.g.*, internet cookies). And with the extraction and collection of data, privacy – for example – become imperilled.

The paradox is clear: on one hand, the efficiency and predictive capacity of algorithms make them a tool for compliance and prevention of offences; on the other hand, this capacity of the machine, driven by big data, raises disturbing alarms linked to a progressive transformation of legal and social systems. This leads to the final question: faced with an AI capable of assessing risks, anticipating harms and acting to prevent them, does it make sense to have a liability system whose categories are built on an event that took place in the past?

These questions, among others, are raised on Part II, under the title "Responsibility".

Anabela Miranda Rodrigues introduces us to the concept of "intelligent corporation" on the chapter "The Last Cocktail – Economic and Financial crime, Corporate Criminal Responsibility and Artificial Intelligence". In a digital economic market, corporations use AI for many purposes, such as business risk assessment, management or monitoring the company. Algorithms are there, making (automated) decisions with a higher degree of autonomy. If legal compliance seems to get more efficient, adequacy problems may rise when confronting

---

[14] Saqib Aziz / Michael Dowling, «Machine Learning and AI for Risk Management», cit., p. 44

[15] Richard A Bales / Katherine Van Wezel Stone, «The Invisible Web at Work: Artificial Intelligence and Electronic Surveillance in the Workplace», *Berkeley Journal of Employment and Labor Law* 41 1 (2020), UCLA School of Law, Public Law Research Paper No. 19-18, SSRN: https://ssrn.com/abstract=3410655

legal models of corporate liability with harms connected to algorithm behaviour. "Still poorly redone from the trapdoor of vicarious responsibility and ambiguities of the organizational defect, finding models of responsibility for corporate crime is, for criminal lawyers, once again urgent".

Miahailis Diamantis, in the Chapter "Algorithmic Harms as Corporate Misconduct", performs a detailed analysis of the existing conceptions of liability, taking it as a premise that "algorithmic harms" do exist. Addressing the algorithmic accountability gap the author is focused on "figuring how to fit algorithms" into liability regimes based on corporate or natural actions. Answering the challenge, the solution – for the present time – is developed around the idea of "beneficial-control account" as criteria for treating algorithmic injuries as corporate actions, covered by corporate law. However, the gap may be open in a disruptive scenario, if the future brings us a world where algorithms are self-performing, self-executing and operate under the control or for benefit of no one. In that case there is no one – corporate or natural – to hold to account.

Christoph Burchard ends the second part of this book with the Chapter "Artificial Intelligence and the End of Criminal Law. On the Algorithmic Transformation of Society", raising disquietening questions about the social and legal transformations created by algorithms. For example, what transformations would result from the introduction of AI applications (predictive policing or "intelligent" sentencing tools) into the system of criminal justice? In the author's own words "what is the status of freedom (especially in a surveillance society needed to power Big Data driven algorithms), trust (especially under the zero trust paradigm that underlies many risk assessment algorithms) and future (especially when algorithms make predictions based on past data) once AI enters into the administration of criminal justice?" These are indeed questions that the criminal law needs to address today "in order to come up with a criminal law that is both (for pragmatic reasons) open to technology as well as (for humane reasons) sensible".

III. Advances in science and technology can be extremely useful in the pursuit of economic efficiency but also of fairness and justice; they

can also be an accelerated path to a securitarian type of law, capable of sacrificing, in a few steps, values conceived as essential in today's society. Some examples may be briefly pointed out, such as the right to privacy and intimacy or the freedom of expression and of choice. Choosing a securitarian law, based on the potential and possibilities that this new technology presents, can have a very high cost in the restriction of fundamental rights by promoting a criminal response to a crime that does not *yet* exist. And with that, a person "labelled" as high-risk by the machine is deprived of the ultimate eventual possibility of not carrying out the (future) crime. A securitarian law, disconnected, in time and space, from a criminal fact (and, therefore, from a real harm to legal values), centred on the agent . It would be a "punitive law" with punishment but no crime to punish, so well portrayed by Alice's doubts in dialogue with the Queen in the beginning of this introduction.

PART I

PREVENTION

1

"Intelligent Compliance"

(https://doi.org/10.47907/livro2021_4c2)

Pedro Maia[1]

Abstract:
Compliance and Artificial Intelligence (AI) are now at the center of banking regulation and banking activity. The way these two realities combine raises a variety of questions, challenging both corporate law and banking law. We try to identify and analyze some of those questions.

Keywords: artificial intelligence; banks; compliance

INTRODUCTION

Ensuring compliance by way of artificial intelligence (AI)[2], which I refer to as "intelligent compliance", is a crossroads of several (r)evolutions which are either underway in the banking sector or which it's keeping track of.

On the one hand, the emergence and growth of a kind of compliance subject to a framework and to a breadth and set of demands

---

[1] Associate Professor of Law. Univ. Coimbra, University of Coimbra Institute for Legal Research, Fac. Law.

[2] The expression appears to first have been used by John McCarthy, in 1956. *See* Scopino, Gregory, *"Key Concepts: Algorithms, Artificial Intelligence, and More"*, in Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and Other Derivatives, Cambridge University Press, Cambridge, 2020, p. 19.

without parallel in the past. Globalization first contributed to this by forcing countries to face added difficulties with regard to the control and prevention of economic crimes, something which shaped the need to call on the banking sector to cooperate in the fight against money laundering and the financing of terrorism ('AML compliance'). This was joined by a new circumstance, which one might say *appeared from within the banking sector itself*, stemming from the regulatory framework come out of the 2007-2008 financial crisis. With the colossal growth of regulatory demands targeted at credit institutions, the re-dimension of the internal system in each of them, so as to ensure the control of and compliance with all the demands imposed on them (regulatory compliance), became imperative.

To this effect, the growth of compliance is a direct reflection of the regulatory structure's huge expansion. The current pandemic now stands side-by-side with the legacy of the financial crisis: until August 2020, over 1330 regulatory measures had been announced by regulators (internationally) and around 15% of prudential regulation was either altered or affected. On April the 2nd, 2020, over 75 publications[3] were made in 24 hours. Technological solutions which allow for the identification of the origin, the classification and the forwarding of regulatory changes to the relevant persons in charge of handling them within a financial institution have become valuable in meeting the demands of regulatory compliance[4].

On the other hand, the growing use of technology, including AI, for compliance resonates both with the function's own capabilities and with the associated risks. "Intelligent compliance" incorporates external risks connected to the technology and to the data used in the function itself.

On another hand still, the emergence of new compliance centralities, whether by turning to non-financial companies for help with the activity (service providers whose scope of business is strictly technological) or by the activity's mentioned re-centering: instead of being centered on knowing the client ('know your client' – KYC), a new

---

[3] *See* JWG, "Out of the window: COVID-19 prompts unexpected regulatory change for 2020 compliance, risk management work plans", 2020 (available at https://www.corlytics.com/newsreleases/out-of-the-window-covid-19-prompts-unexpected-regulatory-change-for-2020-compliance-riskmanagement-workplans).

[4] *See "2021: A Critical Year of RegTech"*, in The Global City, 2021, p. 19.

spotlight shines on data ('know your data' – KYD)[5]. From this evolution, a new stage of *RegTech* will emerge: *RegTech 2.0* will become *RegTech 3.0.*

Problems specific to the banking sector – related to the function which compliance plays therein – are joined by the countless challenges posed by the use of Artificial Intelligence, which the scientific community and authorities are rapidly becoming aware of [6].

First of all, I will present the compliance function by briefly describing its origin, evolution, and current framework. Thereafter, I will succinctly describe the importance of technology to the banking sector in general, after which I will again succinctly present some of the elements necessary to the understanding of AI and, more generally, of automation technologies. A description of the usefulness of such technologies to banking compliance will follow. Lastly, I will reflect on the risks and challenges posed by AI on several different levels.

I. DEFINITION AND EVOLUTION OF COMPLIANCE

The definition of "compliance", in the sense of *observance of the law* (understood in a broad sense) or of "acting in observance of the law", appears at first sight to be nothing more than a truism[7]: the duty to observe the law (in a broad sense) undoubtedly comes from the *principle of the rule of law* and, as such, compliance is neither a recent evolution[8]

---

[5] *See* Jung, John Ho Hee, *"RegTech and SupTech: the future of compliance"*, in FinTech – Law and Regulation, Elgar Financial Law and Practice, United Kingdom, 2019, p. 260, Arner, Douglas W., Barberis, Jànos, and Buckley, Ross P., *"Fintech and Regtech in a Nutshell, and the Future in a Sandbox"*, in CFA Institute Research Foundation, 2017, p. 3, Arner, Douglas W., Barberis, Jànos, and Buckley, Ross P. *"FinTech, RegTech, and the Reconceptualization of Financial Regulation"*, in Northwestern Journal of International Law & Business, Vol. 37, No. 3, 2016, p. 405, Arner, Douglas W, Barberis, Janos Nathan and Buckley, Ross P, *"The emergence of RegTech 2.0: From know your customer to know your data"*, in Journal of Financial Transformation, vol. 44, 2016, p. 7.

[6] *See*, for example, Arner, Douglas W., Barberis, Jànos, and Buckley, Ross P., *"Fintech… "*, cit. p. 6 ff.

[7] In this sense, *see* Uwe Schneider, "Compliance als Aufgabe der Unternehmensleitung", *ZIP*, 2003, p. 646.

[8] In the sense that, as a duty to observe the law, compliance is inherent to the principle of the rule of law, *see Hauschka/Moosmayer/Lösler Corporate Compliance*, 3. Auflage, 2016, annot. 2.

nor possesses its own or specific content. All entities, including those of the banking sector, must therefore observe the law.

But that's neither the current specific meaning of compliance nor the meaning with which it came to be. Indeed, the fact that that's not its meaning is precisely the reason why compliance progressively moved further away from *legal departments*, so as not to be confined to a strict assessment of legal compliance[9].

Compliance may be defined in different ways, holding different characteristics or resulting from different perspectives. It may be defined as a system and set of processes through which an organization undertakes to ensure that its employees and other persons in charge act in accordance with the *"rules"*; besides the law in a strict sense, within these rules one finds the whole regulatory catalogue and the organization's own internal rules such as codes of conduct. Or it may be defined as the "set of internal processes used by a company to *adapt its actions to the applicable rules"*[10]. It may be connected to the "*effort* to ensure that the company and its employees follow legal and regulatory requisites, industry practices, and the company's own policies and internal regulations[11]. Or it may be connected to the "company's set of systems and processes created with the objective of avoiding civil or

---

[9] The advantages and inconveniences of separating compliance from legal services have been highly debated. *See* ARMOUR, JOHN, GARRETT, BRANDON L., GORDON, JEFFREY N. and MIN, GEEYOUNG, *"Board Compliance"*, in Minnesota Law Review, Vol. 104, 2019, p. 1210 ff., and MCNEECE, JOHN B., *"The Ethical Conflicts of the Hybrid General Counsel and Chief Compliance Officer"*, in Georgetown Journal of Legal Ethics, Vol. 25, 2012, p. 677 ff. The matter must be taken into account within the scope of *innkeepers* as *gatekeepers* (but it's a debatable subject: critically, *see* GADINIS, STAVROS and MIAZAD, AMELIA, *"The Hidden Power of Compliance"*, in Minnesota Law Review, Vol. 103, 2019, p. 2154 ff.). On this matter, *see* SIMMONS, OMARI SCOTT and DINNAGE, JAMES D., *"Innkeepers: A Unifying Theory of the In-House Counsel Role"*, in Seton Hall Law Review, Vol. 41, No. 1, 2011, p. 77 ff. (with the eloquent use of the expression "innkeeper" as a reference to persons who act as gatekeepers from within the organization itself).

[10] *See* GRIFFITH, SEAN J., *"Corporate Governance in an Era of Compliance"*, in William & Mary Law Review Online, Vol. 57, No. 6, 2016, p. 2082. In a very similar sense, BAER, MIRIAM HECHLER, *"Governing Corporate Compliance"*, in Boston College Law Review, Vol. 50, 2009, p. 958, OROZCO, DAVID, *"A Systems Theory of Compliance Law"*, in University of Pennsylvania Journal Business Law, Vol. 22, No. 2, 2020, p. 250 ff.

[11] *See* MARTINEZ, VERONICA ROOT, *"The Compliance Process"*, in Indiana Law Journal, Vol. 94, 2019, p. 205.

criminal liability by the organization or its bodies"[12] (italics have been used as a way of highlighting the elements particular to each of the definitions).

Although each of these definitions emphasize different characteristics, none takes on compliance based on the *outcome*: it is not, therefore, about ensuring that *the law is complied with* – including regulatory and recommendatory dispositions and internal regulations, in a very broad sense – but rather about *creating a system* (made up of means, processes, and procedures) with the goal of both avoiding the breach of the legal framework within the company and of ensuring that, should a breach occur, it is detected. Compliance's current theoretic framework is essentially procedural in nature[13], which of course drives it away from a substantial result. Compliance is thus directed at the *prevention of risk* and, because it is so, its worth isn't measured by a case of breach of law (always in a broad sense) that may actually occur, but instead by any breach of law that may *probably* occur in face of the existing system and processes of prevention. The occurrence of a particular breach within the company isn't in and of itself evidence of compliance's fragility – much less of a breach of compliance duties[14]. Conversely, the non-occurrence of a normative breach by itself doesn't mean that no compliance duties have been breached.

Since compliance (much like other control functions in any credit institution) is linked to *risk*, and since a company's resources are limited, the past several years have seen what some authors call a "risk revolution" in internal and external control[15]: the design of internal control systems, including compliance, now consists of a risk evaluation which, after completed, is abided by. This is entirely understandable given that the existing means are finite and must be allocated to areas where a greater risk is detected. This "risk-based approach" has the advantage of allowing the company to essentially focus on the features where there

---

[12] *See* Gunnar Groh, in Creifelds kompakt, Rechtswörterbuch, 4. Auflage, 2021, Beck-online.

[13] *See* Orozco, David, *"A Systems..."*, cit., p. 254 ff. and the very recent *Principles of Law Compliance, Risk and Management, and Enforcement*, of the American Law Institute (§3.01).

[14] Insofar as such a duty exists.

[15] *See* Miller, Geoffrey Parsons, *"Compliance: Past, Present and Future"*, in University of Toledo Law Review, Vol. 48, 2016, p. 446.

is a greater risk of a harmful event occurring, although it's important to acknowledge that the approach itself entails a risk, in that it relies on an inadequate assessment of risk. With that being the case, the systems, which were built on top of a mistake, aren't suitable to prevent the occurrence of a harmful event[16].

This has another highly relevant implication still. A so-called "zero tolerance" to breaches of compliance has repeatedly been heard in the discourse of politicians, regulators, and even regulated entities. This approach is in and of itself *conceptually incompatible* with the officially adopted "risk-based approach". "Zero tolerance" would literally entail something which is unreachable and economically unsustainable: the company being absolutely certain at all times of not being in breach of any rule (in a broad sense) with regard to all of its actions. Such an approach is not only impossible; it would actually be the opposite of a "risk-based approach", which consists exactly of weighting a risk and then determining which issues compliance control should be directed at and which means it should make use of[17].

Compliance also appears to be undertheorized[18]: *compliance law*[19] *is still little studied and little defined as a theoretic unity*, ultimately being determined by somewhat isolated legislative or regulatory interventions and led by practical developments that at any given moment direct its normative content.

---

[16] The path leading to the financial system's sub-prime crisis appears to prove not only a possibly incorrect perception of risk – in general – but also the inability of control systems of preventing that damaging event. *See* Miller, Geoffrey Parsons, "*Compliance...*", cit., p. 447 ff. Another example may surely be found in the (already materialized) risk of a global pandemic, which although possible was not identified.

[17] *See* Miller, Geoffrey P., *"Risk Management and Compliance in Banks: The United States and Europe"*, in European Banking Union, Oxford, United Kingdom, 2015, p. 211.

[18] *See* Griffith, Sean J., *"Corporate..."*, cit., p. 2081, and Orozco, David, *"A Systems..."*, cit., p. 246.

[19] We will not delve deeper into the hotly debated issue of knowing whether compliance is an independent field of study. *See*, for example, Sokol, D. Daniel, *"Twenty-Eighth Annual Corporate Law Center Symposium: Rethinking Compliance"*, in University of Cincinnati Law Review, *V*ol. 84, No. 2, 2016, p. 401 ff. (highlighting the huge variety of understandings when it comes to compliance and the resulting difficulty in creating a field of law), Martinez, Veronica Root, *"The Compliance..."*, cit., p. 244, and Orozco, David, *"A Systems..."*, cit., p. 251 ff.

Compliance's somewhat theoretic vagueness may be attributed to its origin, wherein two distinct paths of evolution can be found: one of a *practical* and *managerial* nature, dictated by the convenience of creating a specific function for internal control independent from legal departments; another of a *regulatory* or *legislative* nature, dictated by the (legislators' and regulators') need to introduce within organizations a body meant to either ensure the observance of the applicable normative structure or prevent transgressions within the company. The first corresponds to what may be referred to as compliance's *positive side*, in which it acts as an instrument or element which strengthens the business and allows for its success; the second corresponds to compliance's *negative side*[20]*, in which it serves the purpose of avoiding or preventing the organization from breaching its legal background. This negative side* may in turn take very different characteristics depending on the regulators' approach: it can be more *prescriptive*, imposing contents *specific* to internal control on the entities supervised; or it can be more *flexible*, granting companies ample freedom in deciding their own systems.

The first mentioned path of evolution is guided by the company's interests and, because it is developed from a judgement of *opportunity* and *convenience of management*, leaves compliance subject to the *management's discretion* in light of the interests pursued by the company and, more important, of its shareholders. In this path, compliance is also an instrument destined to satisfy the interests pursued by the company and is thus *in line with* one view of corporate interest – coinciding with that of the shareholders (profit or maximization of value), should that be the case. From this perspective, compliance is, after all, the *management of corporate risk*[21] *– in this case, the risk of breaching the*

---

[20] Also making this distinction, *see* Cunningham, Lawrence A., *"The Appeal and Limits of Internal Controls to Fight Fraud, Terrorism, Other Ills"*, in The Journal of Corporation Law, Vol. 29, 2004, p. 267 ff., and Chiu, Iris H-Y, *"Regulating (From) the Inside. The Legal Framework for Internal Control in Banks and Financial Institutions"*, Hart Publishing, Oxford, 2015, p. 8 ff.

[21] The management of corporate risk may be defined as the process through which the management body delineates the strategy and objectives that will allow the company to reach an optimal balance between growth, return, and related risks. *See* Bainbridge, Stephen M., *"Caremark and Enterprise Risk Management"*, in The Journal of Corporation Law, Vol. 34, 2008, p. 967. In a similar sense, *see* Der Elst, Christoph and Van Daelen, Marijn, *"Risk Management in European and American Corporate Law"*, in ECGI-Law Working Paper, No. 122, 2009, p. 6.

law and of having to face the consequences arising therefrom – and ends up overlaying or falling within so-called risk management: the system designed to handle all risks which a company is exposed to[22].

The second path of evolution, dictated by legislators and regulators and appearing at a later stage, most notably after the 2007-2008 financial crisis, is of a completely diverse nature. It's not about compliance as an instrument aimed at the pursuit of corporate interests, but instead as a safeguarded set of (legal and regulatory) dictates: a way of *ensuring* that the company's business does not harm the interests that such dictates seek to protect; interests which naturally do not coincide with those of the company but (as well) with those of *third parties*, with *public interest*, with the interests of certain categories of persons[23]. In this second path of evolution compliance is no longer an *instrument in the satisfaction of corporate interest* – therefore of a discretionary nature, defined and limited by each company's freedom in management – but instead an instrument designed to satisfy interests foreign and unavailable to the company – therefore of an *imperative* and *hetero-determined* nature[24].

It truth, besides these two paths of evolution, a third, more visible in jurisdictions such as the United States of America, may still be identified. In it, compliance plays a rather indirect and instrumental role, although still with great practical relevance with regard to one point in particular: that of the *accountability*, above all criminal, of

---

[22] *See* Bainbridge, Stephen M., *"Caremark..."*, cit., p. 968 (defending that between risk management and compliance there is no difference of nature, only a difference of level).

[23] Defending the dimension of social responsability, *see* Rodrigues, Anabela Miranda, *Direito Penal Económico: Uma Política Criminal na Era Compliance*, 2nd Ed. Almedina, Coimbra, 2021, p. 91 ff. For a different understanding (compliance as a function *of the company* and *for the company*), although much earlier than the function's recent evolution, *see* Labareda, João, *"Contributo para o estudo do sistema de controlo e da função de cumprimento ("Compliance")"*, in Direito dos Valores Mobiliários, 2016, p. 364.

[24] *See* Lösler, Thomas, *"Das moderne Verständnis von Compliance im Finanzmarktrecht"*, in NZG, 2005, p. 106, Weber-Rey, Daniela, *"Der Aufsichtsrat in der europäischen Perspektive – Vorschläge und Ideen für eine wirksame Corporate Governance"*, in NZG, 2013, p. 766 (which even refers that the evolution came at the cost of "corporate freedom"), Gebauer/Niermann, in *Hauschka/Moosmayer/Lösler...,* cit., § 48, annot. 19, and Maia, Pedro, *"Direito das Sociedades Bancárias"*, in Revista de Legislação e de Jurisprudência, Year 149, No. 4023, 2020, p. 398.

company directors. Starting in the 1990's, the existence of a compliance function within companies began being taken into account for the purposes of criminal, or even civil, liability. Some authors even identify 1991's *Sentencing Guidelines for Organizations* as the beginning of the current stage of compliance, in that they represent the first indicators of the relevance attributed to the existence of an "effective compliance program" within companies in reducing penalties[25].

Case law[26] soon followed by recognizing the existence of a duty to implement a reports and information system by the company's management body. And should the system signal a problem – a so-called "red flag" – the management body must act in a way that gathers the facts and takes the appropriate measures. It's important to underline that although public intervention left a mark of its influence (particularly when it comes to criminal prosecution), in this path of evolution the state neither *imposed* nor *determined* the existence of corporate programs of compliance. A program was not seen as a company's *legal duty*, despite an advantage – an indirect incentive – being offered by its implementation: the benefits which would come to the company and its directors should an event give rise to liability. These were therefore "explicit incentives" given by the state to the implementation of compliance programs seen as mitigating factors in the sentencing of corporations[27].

---

[25] *See* Griffith, Sean J., *"Corporate..."*, cit., p. 2084, Hess, David, *"Ethical Infrastructure and Evidence-Based Corporate Compliance and Ethics Programs: Policy Implications from the Empirical Evidence"*, in New York University Journal of Law and Business, Vol. 12, 2015, p. 318, and Langevoort, Donald C., *"Cultures of compliance"*, in American Criminal Law Review, Vol. 54, 2017, p. 940 ff., Garrett, Brandon L. and Mitchell, Gregory, *"Testing Compliance"*, in Law and Contemporary Problems, Vol. 83, No. 4, 2020, p. 49 ff. Amongst ourselves, *see* Rodrigues, Anabela Miranda, *Direito...* cit., p. 116 ff., and Sousa, Susana Aires de, *"A colaboração processual dos entes coletivos: legalidade, oportunidade ou "troca de favores"?"*, in Revista do Ministério Público, n.º 158, 2019, pp. 9ff. (with an important assessment of the evolution and of its implications for penal law and penal procedure). The reduction of sentencing due to the existence of an effective compliance program could be as far as 95% (*see* Gadinis, Stavros and Miazad, Amelia, *"The Hidden..."*, cit., p. 2146).

[26] In the 1996 case *In re Caremark Int'l Inc. Derivative Litig.*, tried in Delaware.

[27] *See* Armour, John, Garrett, Brandon L., Gordon, Jeffrey N. and Min, Geeyoung, *"Board..."*, cit., p. 1195, Gadinis, Stavros and Miazad, Amelia, *"The Hidden..."*, cit., p. 2148 ff.

This evolution was made complete by the United States Department of Justice's guidelines regarding the relevance of "effective" programs of compliance[28] in the potential prosecution of companies. And in the first years of the new millennium, in the midst of new frauds and scandals of accounting and auditing, the *Brooklyn Plan* was set in motion: in exchange for non-prosecution agreements, companies would pay penalties and fines and adopt rigorous programs of compliance[29]. It was in the context of these agreements of non-prosecution or of deferred prosecution[30] – the effects of which have been highly criticized[31] – that it became common to demand companies to implement programs of compliance typically centered on the approval of policies and processes directed at employees subject to training and monitoring[32].

---

[28] A matter which I will not delve into has been a special subject of debate: that of knowing which requisites are necessary to consider a compliance program "effective". This matter is very relevant because it's about knowing if the program's effectiveness is assessed by its *result* – by its efficiency – or solely by its *structure* and allocated *means*. Some authors point the risk (or even fact) that some compliance programs may become nothing more than "box-ticking" exercises – a simple demonstration that a compliance program exists – wherefrom the advantages expected from the organization's effective compliance and from a culture supportive of it did not result, or may not have resulted. This even justifies calling such programs "always elusive", or evasive (the origin of this expression is Martinez, Veronica Root, *"The Compliance..."*, cit., p. 205). On this matter, *see*, for example, Langevoort, Donald C., *"Monitoring: the behavioral economics of inducing agents' compliance with legal rules"*, in Georgetown University Law Center Business, Economics and Regulatory Policy, Law and Economics Research Paper, No. 276121, 2001, p. 933 ff., Armour, John, Gordon, Jeffrey and Min, Geeyoung, *"Taking Compliance Seriously"*, in Yale Journal on Regulation, Vol. 37, No. 1, 2020, p. 15 ff., Griffith, Sean J., *"Corporate..."*, cit., p. 2105 ff. (the metrics on evaluating the effectiveness take into account the *activity* instead of the *impact*, "showing that compliance should be busy but not necessarily effective"), Gadinis, Stavros and Miazad, Amelia, *"The Hidden..."*, cit., p. 2139, and Garrett, Brandon L. and Mitchell, Gregory, *"Testing..."*, cit., p. 56 ff.

[29] In this regard, *see* Garrett, Brandon L, *Too Big to Jail*, Harvard University Press, Cambridge, 2014, p. 54 ff. (which establishes a connection between the evolution of compliance and the criminal investigation of companies).

[30] *Deferred Prosecution Agreements* ('DPA') and *Non-Prosecution Agreements* ('NPA').

[31] *See* Langevoort, Donald C, *"Cultures..."*, cit., p. 970 ff.

[32] *See* Griffith, Sean J., *"Corporate..."*, cit., p. 2088 ff.

Although the rise of compliance as a sectorial regulatory reality had already occurred before[33], the determining factor in its *significant progress* was the 2007-2008 financial crisis: the relevant regulatory framework had been "bare-boned" until then[34]. After identifying the breach of credit institutions' internal policies – governance rules[35] – as the explicit cause of the crisis, supervisors (and legislators) moved decisively forward and *imposed* specific compliance duties to the financial sector. Which may define the new framework of compliance has a "reactive process", determined by the occurrence of scandals and crimes which propel legislators and legislators to intervene[36]. The legal and regulatory framework of this new outlook

---

[33] In April, 2005, the Basel Committee on Banking Supervision published it's report titled "Compliance and the compliance function" and, also in that year, the Bank of Portugal published Instruction 20/2005, which amended Instruction 72/96 by expressly pointing out the risk of compliance. Curiously, that risk was then inserted in "risk management", where it was defined as "the risk of an institution being subject to legal or regulatory sanctions or financial or reputational losses as a result of not having abided by the laws, norms, codes of conduct, or standards of "good practice" – as may be read in the Instruction's introduction. In this regard and on compliance's progressive reception by the Portuguese regulatory system, *see* Labareda, João, *"Contributo..."*, cit., p. 296 ff. and, more recently, Bastos, Nuno Moraes, *"Corporate Governance, Compliance e a Função Compliance nos Setores Bancários e Segurador"*, in A Emergência e o Futuro do Corporate Governance em Portugal, Vol. II, Almedina, Coimbra, 2018, p. 207 ff.

[34] The expression is from Chiu, Iris H-Y, *"Regulating…"*, cit., p. 6.

[35] This is a controversial matter where two theories collide: the "theory of irrelevance", which doesn't see failures in governance as the origin of the crisis, and the "theory of *force majeure*", according to which those failures are the crisis' major cause. The right position seems to be recognizing that although governance was *one* of the key factors of the crisis, it was not the *determining* factor, or even *the most important*. *See* Maia, Pedro, *"Direito..."*, cit., p. 379 (and the bibliography referred therein).

[36] In this regard, *see* Orozco, David, *"A Systems..."*, cit., p. 254 ff. But the quality of this approach's result is highly debatable and, on the plane of theoretical analysis itself, highly open to criticism, especially due to the fact that it ignores the influence the social and economic context has in the behavior of individuals, organizations, and institutions as a determining factor of compliance's result. *See* Orozco, David, *"A Systems..."*, cit., p. 257 ff. For an analysis of the issue of legislation passed as a reaction to crises and scandals (in the words of Arner, Douglas W., Barberis, Janos Nathan and Buckley, Ross P., *"The emergence of RegTech 2.0: From know your customer to know your data"*, cit., p. 8, "the history of global financial institutions is the story of regulatory initiatives in response to crisis"), *see* Banner, Stuart, *"What causes new*

of compliance – or better still, of this new nature of compliance and new connection to corporate governance – arrived as part of the "legislative tsunami" or "regulatory deluge" that the 2007-2008 financial crisis unleashed[37]. In European and Portuguese law, one should highlight Directive 2013/36/EU (known as 'CRD IV'[38]) and the accompanying Regulation (EU) No. 575/2013 (known as 'CRR'[39]). Though CRD IV practically doesn't address the issue, with the exception of an indirect reference to "compliance functions" in Article 92, Paragraph 2, Section f) of the directive's Portuguese version, the basis for the regulation of internal control and for an intervention by the EBA are set therein (*see* Article 74, Paragraph 1) – an intervention which at any rate had already taken place in 2011, with the publication of the *Guidelines on Internal Governance* ('GL 44'[40]), where the existence of an *autonomous* internal control function – the compliance function – which may only be combined with the risk management function in smaller or less complex institutions (*see* Paragraph 24.5[41]), is determined. The compliance function is regulated

---

securities regulation? 300 years of evidence", in Washington University Law Quarterly, 75, No. 2, 1997, p. 849 ff., Coffee, John C. Jr., *"Political Economy of Dodd-Frank: Why Financial Reform Tends to be Frustrated and Systemic Risk Perpetuated"*, in Cornell Law Review, Vol. 97, No. 5, 2011, p. 1020 ff. (who identifies the regulation of the financial system as a "sine curve" – a repetitive and soft oscillation).

37 *See* Maia, Pedro, *"Direito..."*, cit., p. 372 (where an annotation containing a description of the most important normative instruments on which that tsunami was based can be found).

38 Amended by Directive (EU) 2019/878 of the European Parliament and of the Council, of May 20th, 2019 (sometimes referred to as 'CRD V'), in the meanwhile.

39 Amended by Regulation (EU) 2019/876 of the European Parliament and of the Council, of May 20th, 2019 (referred to as 'CRR II').

40 It's important to clarify that the EBA's *Guidelines*, although apparently nothing more than recommendatory soft law, end up representing what some authors call "hoft law", in the sense that they appear to be soft law when issued but turn into hard law when national regulatory supervisors convert the recommendations therein into orders which regulated entities are subject to. In this regard, *see* Maia, Pedro, *"Direito..."*, cit., p. 400.

41 The fact that compliance might sometimes not be autonomous at the organizational plane explains why the legislator and the European regulators do not refer to a "compliance department" but to a "compliance function": the latter is *mandatory*, without any exceptions, but it's assignment to an *autonomous department* is not. In this regard, *see* Gebauer/Niermann, *"Hauschka/Moosmayer/Lösler..."*, cit., p. 22, § 48, annot. 6.

thereafter (*see* Paragraph 28 and following). A new version of the *Guidelines* was published in 2018[42].

In a way, this evolution represents a veritable *transmutation* of compliance, which, no longer confined to the company's circle of *autonomy of* (risk) *management*, becomes (at least to some extent) part of the domain of legislative or regulatory intervention. While appealing to variable terms and distinct measures, legislators and regulators imposed on financial sector companies the duty of setting up an (internal) compliance function. As mentioned before, the development of compliance had already received *external boosts*, but now its existence became *externally determined*. Though developed and secured *internally* – one must not forget that compliance is an *internal* control function –, it presently has an *exogenous* origin when it comes to banking companies, in the sense that it took from the management body the freedom not only to decide on its existence, but also on multiple aspects of its structure and operation[43]. It's the legislator and the regulator who determine them. This governance is therefore *internal* to the company but imposed on it by *external* sources[44].

While needfully brief and even incomplete, the framework presented above allows the understanding of the new context which compliance is a part of within baking sector companies. A function of *internal* control which, while taking place *within the company*, serves purposes that are not exclusively inherent to the company itself when understood as an instrument at the service of shareholder interests[45].

---

[42] *Guidelines on Internal Governance* (EBA/GL/2017/11, of March 21st, 2018, available at https://www.eba.europa.eu/sites/default/documents/files/documents/10180/2164689/151a6ca3-31ae-40b0-9f55-9d6c65b86b00/Guidelines%20on%20Internal%20Governance%20%28EBA-GL-2017-11%29_PT.pdf?retry=1). In this regard, amongst ourselves, *see* Fonseca, Patrícia Afonso, *"As Novas Orientações da EBA em Matéria de Governo Interno"*, in A Emergência e o Futuro do Corporate Governance em Portugal, Vol. II, Almedina, Coimbra, 2018, p. 235 ff.

[43] Highlighting the *exogenous* origin of compliance, which contrasts with the function's *internal* nature, *see* Griffith, Sean J., *"Corporate..."*, cit., p. 2078 ff.

[44] *See* Griffith, Sean J., *"Corporate..."*, cit., p. 2079 ff.

[45] What I stated above does not contend with the heated discussion which has been taking place amongst authors (and even amongst the public) on the issue of companies' purpose – do they follow their shareholders' selfish interests or others beyond that? If so, which ones and on which terms? –, a discussion stimulated by the "corporate purpose" current of thought. Should one follow this tendency there will be some facets of compliance found to overstep a company's corporate purpose. On

II. TECHNOLOGY IN BANKING

Banking has always been particularly open to technical innovation and progress[46]. In some cases, instead of merely *accepting* this innovation and progress, it went so far as *promoting* it (as *creating* it, in a sense). One need only think of the *telegraph*, introduced in 1838 and promptly incorporated in the daily activity of banks. And of the first *transatlantic cable*, laid in 1866 and soon after already facilitating intense financial exchanges between Europe and the United States of America – and driving the first globalization of financial activity at the end of the nineteenth century, through the rapid transmission of information, transactions and payments. In 1958, *Bank of America* and *American Express* introduced the *credit card*, a technology-based revolution in lending and payment systems.

In 1964, *Xerox* introduced the first commercial fax machine (under the name *Long Distance Xerography*, or 'LDX'), which would become widely used in the financial sector; in 1966, a global telex network that ensured the quickness and safeness of communications in financial transactions was already in place.

In 1967, *Barclays Bank* introduced a ground-breaking system of automatic cash withdrawal and money transfer – the *Automatic Teller Machine*, or 'ATM' –, one of the most consequent technology-based revolutions in banking until the present day. *Calculators*, invented by *Texas Instruments* also in 1967, were immediately adopted by the sector.

This stage, which came to an end in the 1960's and may be called *FinTech 1.0*, rested on *analogical* technology. What followed was a

---

the matter of "corporate purpose" and the intense debate surrounding it, *see*, with particular relevance and disagreeing positions, Mayer, Colin, *"The future of the corporation: Towards humane business"*, in Journal of the British Academy, Vol. 6, No. 1, 2018, p. 1 ff., Bebchuk, Lucian A. and Tallarita, Roberto, *"The Illusory Promise of Stakeholder Governance"*, in Paper SSRN, 2020, p. 1 ff., Rock, Edward B., *"For Whom is the Corporation Managed in 2020?: The Debate over Corporate Purpose"*, in European Corporate Governance Institute – Law Working Paper, No. 515, 2020, p. 1 ff., and Lipshaw, Jeffrey M., *"The False Dichotomy of Corporate Governance Platitudes"*, in The Journal of Corporation Law, Vol. 46, No. 2, 2021, p. 346 ff.

[46] This is stated by the European Commission in its *FinTech Action Plan: For a more competitive and innovative European financial sector*, 2018, p. 2.

shift to *digital* technology until the late 1980's, intensified by the crash of the New York Stock Exchange in 1987 – a stage which some authors identify as *FinTech 2.0*. With the development of the *World Wide Web* in the 1990's, the first online banking service was launched by the North American bank *Wells Fargo*. The first online banks without traditional brick-and-mortar branches, such as *ING Direct* or *HSBC Direct*, appeared in 2005[47].

This very brief historical overview of the development of technology in banking helps to understand that the technological evolution brought about by Robotics and AI isn't in and of itself an *irregular*, *strange* or even *novel* situation in the industry: financial activity has always promoted and surrounded itself with the most developed tools and instruments that technology has to offer at each point in time[48].

Although the incorporation of new technical or technological means in the financial business isn't a novelty, the current situation is new mostly because of *two aspects*[49]. *The first of these concerns is the fact that new technologies, which are undoubtedly being assimilated by companies within the sector, are mostly used by non-financial companies* – or companies not financial in nature. These aren't financial companies taking advantage of a *new technology* to conduct their *old trade*; in most cases, they're companies technological in nature taking advantage of technology (already existent to them) to conduct a *new trade*. *FinTech*

---

[47] An historical overview of the financial sector's technological evolution can be read, for example, in Arner, Douglas W., Buckley, Ross P. and Barberis, Janos N., *"The Evolution of Fintech: A New Post-Crisis Paradigm?"*, in Georgetown Journal of International Affairs, Vol. 47, 2016, p. 1274 ff., and in Jung, John Ho Hee, *"RegTech..."*, cit., p. 257 ff.

[48] It need only be said that *Goldman Sachs* employs 33 thousand engineers, more than those employed by *Twitter*, *Facebook*, or *LinkedIn*, something that is quite revealing of the technological level already reached by the banking sector. *See* Arner, Douglas W., Buckley, Ross P. and Barberis, Janos N., *"The Evolution..."*, cit., p. 1291. Or that *JP Morgan Chase* is estimated to have more software developers than *Google* or *Microsoft* (*see* Lin, Tom C. W., *"Compliance, Technology, and Modern Finance"*, in Brook. J. Corp. Fin. & Com. L., Vol. 11, 2016, p. 161).

[49] The fact that the financial sector has always adopted technical innovations so quickly does not mean that it's quick to receive "technological disruptions", as is the case. In the sense that the financial sector has always resisted and suspected disruptive innovations, *see* Anagnostopoulos, Ioannis, *"Fintech and regtech: Impact on regulators and banks"*, in Journal of Economics and Business, Vol. 100, 2018, p. 11.

and *TechFin* companies, to those who know the difference, are precisely that[50].

As it's been frequently highlighted, technological evolution is *opening up* the financial sector – *opening up* also in the sense of *freeing* the activity, at least temporarily, because the traditional legal framework isn't capable of regulating and supervising these new forms of financial activity. These so-called *FinTech* companies – *Fin* (Financial) + *Tech* (Technology), which consists of using technology to provide all manner of financial services[51] – under many ways escape the existing legal and regulatory framework. And what's more, despite technology being what *operatively* supports them, it's the *legal framework* which at least partially stimulates them *economically*. As a matter of fact, the activity's boom after the financial crisis is no *mere coincidence*: the great crisis fostered a significant reinforcement of the regulatory framework and consequently occasioned an equally significant rise in the associated costs incurred in by companies having to comply with it, so that conducting the activity "absent of regulation" became a major competitive advantage[52].

This represents a very relevant profile for the analysis and debate of technological evolution: in which way it should be made a part of the regulatory framework, should that framework be shared or separated, how should the regulatory entities themselves evolve, and how can they be made capable of handling these new phenomena[53]. This

---

[50] On the matter of *FinTechs*, among an extensive bibliography but discussing some conceptual aspects only, *see* Bradley, Christopher G., *"Fintech's Double Edges"*, in Chicago-Kent Law Review, Vol. 93, No. 1, 2018, p. 77 ff., Brummer, Chris and Yadav, Yesha, *"Fintech and the Innovation Trilemma"*, in The Georgetown Law Journal, Vol. 107, 2019, p. 241, annot. 18, and Baumanns, Charlotte, *"Fintech als Anlageberater? Die aufsichtsrechtliche Einordnung von Robo-Advisory"*, in BKR, 2016, p. 366 ff.

[51] *See*, for example, Arner, Douglas W., Buckley, Ross P. and Barberis, Janos N., *"The Evolution..."*, cit., p. 1272.

[52] *See* Arner, Douglas W., Buckley, Ross P. and Barberis, Janos N., *"The Evolution..."*, cit., p. 1286. The history of the financial system's regulation and of its tendencies and interactions must inform the decisions that require in response to new tendencies. For a history from this perspective, *see* Marco, Lamandini and Munoz David, Ramos, *"A brief history of the evolution of financial institutions and of their regulation"*, in EU Financial Law. An Introduction, Cedam, Padova, 2016, p. 3 ff.

[53] On this matter, *see*, for example and among many others, Fein, Melanie L., *"How Should Robo-Advisors Be Regulated? Unanswered Regulatory Questions"*, in Allianz Global Investors, 2017, p. 1 ff.

naturally comes in addition to the assessment of the economic and social impacts which the adoption of these new technologies entails at various levels: the reduction of financial companies' operating costs, the democratization of services (allowing them to reach sections of the population where resources are not as available, although with that favoring a better allocation of significantly valued economic resources), the improvement of investment decisions (based on more rationally processed and technically capable information), the increase of market efficiency, etc.[54] To some, the length and depth of what is called the financial industry's "technological revolution" commands the phenomenon's analysis in a way that's not merely micro-transactional but also systemic, due to the fact that its impacts have even been felt at the level of politics and power relations[55]; to this, the realization that "software eats the world", i.e. that it subjugates all other industries – the financial services industry is but one example – and forces their total reconversion[56], must be added. In the 1940's, Schumpeter theorized about the gale of "creative destruction" in the economy[57]: regardless of the theory's correctness, the concept may surely be used to illustrate the implications associated with the use of software (including robotics and AI) in the financial industry.

This is not, however, the object of this study.

The other feature where the situation is new concerns the *speed* with which the evolution is happening[58]. And one must not think that this is purely related to time and in no way relevant beyond that;

---

[54] On these implications, *see*, for example, Lin, Tom C. W., *"Artificial Intelligence, Finance, and the Law"*, in Fordham Law Review, Vol. 88, 2019, p. 531 ff. (especially highlighting the assessment and analysis of the risks and dangers inherent to the use of robotics and AI by financial services).

[55] *See* Omarova, Saule T., *"New Tech v. New Deal: Fintech as a Systemic Phenomenon"*, in Yale Journal on Regulation, Vol. 36, 2019, p. 735 ff.

[56] The expression belongs to Marc Andreessen, "Why software is eating the world", in *Wall Street Journal* (20.08.2011).

[57] *See* Schumpeter, Joseph, *Capitalismo, Socialismo e Democracia*, Actual Editora, Coimbra, 2018, p. 119 ff. Although the expression most recently used is "disruption" or "disruptive effect" (for example, Piri, Michael M., *"The Changing Landscapes of FinTech and RegTech: Why the United States Should Create a Federal Regulatory Sandbox"*, in Business & Finance Law Review, Vol. 2, No. 2, 2019, p. 236), the general meaning remains the same.

[58] *See* Arner, Douglas W., Buckley, Ross P. and Barberis, Janos N., *"The Evolution..."*, cit., p. 1276.

evolution at a very rapid pace itself represents an *increased risk* for incumbent companies, challenged (competitively *attacked*, strictly speaking) by new players which themselves pose several other risks: of companies failing in the face of competition – thus compromising the stability of the financial sector; of rigid and inadequate legal output, incapable of handling new phenomena; or of legal output which, faced with the need to respond quickly to new situations, may be rushed and inconsistent and thus give way to undesirable consequences[59].

III. ARTIFICIAL INTELLIGENCE AND *REGTECH*

There is no consensual and widely accepted definition of AI[60]. For the purposes of this study, the definition used in the European Commission's proposal for an Artificial Intelligence Act[61] (Article 3, Paragraph 1), issued on April, 2021, will be adopted: "[an] 'artificial intelligence system' (AI system) [is a] software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with". In turn, the proposal's Annex I identifies the following AI techniques and approaches: "(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference

---

[59] An interesting analysis resting on the understanding that the evolution brought about by *FinTechs* differs from the ones preceding may be read in Brummer, Chris and Yadav, Yesha, *"Fintech..."*, cit., p. 242 ff.

[60] *See* Scopino, Gregory, *"Key...",* cit., p. 19, and Yang, Yueh-Ping (Alex) and Tsang, Chengyun, *"RegTech and the New Era of Financial Regulators: Envisaging More Public-Private-Partnership Models of Financial Regulators"*, in University of Pennsylvania Journal of Business Law, 2018, p. 363 ff., where two different definitions, corresponding to two different visions, are confronted: one which connects *RegTech* to the technologies which facilitate *communication* between regulators and regulated entities; another which connects it to the *development of the regulatory system*.

[61] Available at https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52021PC0206.

and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods"[62].

AI itself isn't something new – it was first referred to in 1956 and effectively developed in the 1970's; but the pace at which it has evolved recently is unprecedented. A confluence of factors helped this radical acceleration: the extraordinary growth of *data* accessible by computer[63] – to which the massive use of internet was decisive, leading some to say that "digitalization is everything"[64]; its *storage* – through the development of clouds which enable the storage of colossal amounts

---

[62] The definition used in the proposal is based on studies promoted by the European Commission with regard to this matter. *See* the *High Level Expert Group on Artificial Intelligence* ("A definition of AI: Main capabilities and scientific disciplines") (available at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341), of 2019, where the following definition was proposed (p. 6): "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)." At an institutional level, *see* the 2018 *OECD Council Recommendation on Artificial Intelligence* (available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449), adopted by the G20 in 2019 (available at https://www.mofa.go.jp/files/000486596.pdf).

[63] The European Commission estimates that 175 zettabytes of data (over five times more than the 33 zettabytes of data produced in 2018) will be produced in 2025. *See "Livro Branco sobre a inteligência artificial – Uma abordagem europeia virada para a excelência e a confiança"*, in European Commission, 2020, p. 4. A zettabyte corresponds to 1 trillion (1.000.000.000.000) gigabytes.

[64] *See* Buckley, Ross P., Zetzsche, Dirk A., Arner, Douglas W. and Tang, Brian W., *"Regulating Artificial Intelligence in Finance: Putting the Human in the Loop"*, in Sydney Law Review, Vol. 43, No. 1, 2021, p. 46, quoting Schwab, in respect of a "Fourth Industrial Revolution". There is even talk of the emergence of a "data economy", an activity of great value consisting in the collection and monetization of data. In this regard, *see* Magnuson, William, *"A Unified Theory of Data"*, in Harvard Journal on Legislation, Vol. 58, 2021, p. 24 ff.

of information at a very low cost[65]; *communication* – data exists in and flows through computers, smartphones, social networks, search engines, etc., widely used all around the world; and *computing power* – according to Moore's law, the number of transistors in a microchip doubles every two years[66], to the point where *quantum computing* is already under way.

The arrival of AI not only allowed persons to be replaced when performing certain tasks, but also made available services that persons would never be able to provide, no matter how many of them or how well prepared they might have been. Therefore, it's not about replacing persons by performing tasks *exactly how they would perform them* – although surely quicker, with less variations in quality and with less mistakes –, but about providing a service which *exceeds human capacity*. AI not only surpasses a human person in *how* – first and foremost with regard to speed –, but in *what*, the end result of the activity. In its current stage of development, AI already offers a wide array of uses[67].

In 2015, the term *RegTech* first appeared, used by Philippe Treleaven[68] and defined by the *Financial Stability Board* ('FBS') as a subset of *FinTech* corresponding to technologies which may facilitate compliance with regulatory demands in a more efficient and effective way than allowed by existing capacities[69]. Still, *RegTech* is not always a part of *FinTech* – that is, it isn't necessarily a part of the latter and therefore is not one of its subsets – because, unlike *RegTech*, it entails a disruptive use of technology. *RegTech* helps companies (whether they are *FinTech* companies or not) comply with regulatory demands through the use

---

[65] According to "Kryder's Law", the quality and capacity of data storage has drastically increased while at the same time costs have decreased, meaning there has been a constant growth in the volume of data collected and stored. *See* Buckley, Ross P., Zetzsche, Dirk A., Arner, Douglas W. and Tang, Brian W., *"Regulating..."*, cit., p. 46.

[66] A comprehensive account of the reasons which propelled AI's evolution can be read in Buckley, Ross P., Zetzsche, Dirk A., Arner, Douglas W. and Tang, Brian W., *"Regulating..."*, cit., p. 46.

[67] *See* the very significant data gathered by the *Bank of England* and the FCA in *Machine learning in UK financial services*, 2019, p. 8 ff.

[68] *See* Treleaven, Philip, *"Financial regulation of FinTech"*, in Journal of Financial Perspectives, 3, 2015, p. 114 ff.

[69] *See* Authority, Financial Conduct, *"Call for Input: Supporting the development and adoption of RegTech"*, available at https://www.fca.org.uk/publication/call-for-input/regtech-call-for-input.pdf, 2015.

of technology. In this sense, *RegTech* and *FinTech* differ in their origin, goals and scope[70].

But the use of Technology as an instrument of compliance came much earlier than the emergence of *RegTech*. The increasing regulatory demands and above all the prevailing regulatory model had already occasioned a growing use of technology, endorsed by the regulators themselves.

To understand the relevance that the regulatory model may have in the use of technology it must be kept present that regulation may target one of three levels of activity of the regulated entity: *planning*, *performance* (action), or *result* (whether positive or negative)[71]. When it targets the *result* (which corresponds to the *"performance-based"* model) the regulator will set rules imposing a certain result. Contrarily, if instead it targets the *performance* (action) the regulator will set rules imposing the use of specific technologies or behaviors to be followed by the regulated entity when performing its activity (*"technology-based"* models).

In turn, the so-called "process-based" or "management-based" model (the latter expression belonging to Cary Coglianese and David Lazer)[72] is characterized by imposing on regulated entities the *flexible* fulfillment of *public interest objectives*, while granting them the freedom (but also the responsibility) to create plans which, in light of the specific information available to them about their own organization, allow them to reach the targets set by the regulator[73]. Thus, risk, which is contextual and expresses itself differently in heterogeneous companies, may be more adequately mitigated by decisions made by each regulated

---

[70] *See* Arner, Douglas W., Barberis, Janos and Buckey, Ross P., *"FinTech..."*, cit., p. 371.

[71] In this regard, *see* Coglianese, Cary and Lazer, David, *"Management-based regulation: Prescribing private management to achieve public goals"*, in Law & Society Review, 37, 4, 2003, p. 693 ff.

[72] *See* Coglianese, Cary and Lazer, David, *"Management-based regulation: Prescribing private management to achieve public goals"*, cit., p. 692 ff. (highlighting that other authors have used distinct expressions to refer to understandings close to each other in meaning, such as "enforced self-regulation", "mandated self-regulation", "reflexive regulation", or "process-based regulation"). The expression "management--based" has a wider scope since it includes a group of processes, systems, and internal management policies that the regulator demands from regulated entities.

[73] *See* Coglianese, Cary and Lazer, David, *"Management-based regulation: Prescribing private management to achieve public goals"*, cit., p. 694 ff.

entity with the aid of the specific information available to them about themselves – instead of by imperative rules uniformly and generally dictated by the regulator[74]: the regulator doesn't determine in which way the regulated entity should comply but instead demands that it set up its own compliance systems and prove that these are adequate to the fulfillment of the objectives[75]. It may be added that, regardless of its theoretic merits, this approach is a *practical inevitability* – regulatory compliance rests on the regulated entity's systems and can't be guaranteed by the regulator –, so that in the end it's about consciously recognizing this reality as an element of the regulator's strategy[76].

A so-called "meta-regulation"[77] or "regulation of self-regulation"[78] was thus born: the regulator creates a general, not too prescriptive outline of a structure and sets certain objectives which must be reached. In turn, the regulated entity keeps its discretion when choosing how to implement the systems and processes necessary to reach the relevant objectives. The regulator intervenes only at a "meta-level", which consists of evaluating plans and subsequently verifying that the regulated entity has followed the plans that it has created itself.

In the field of finance, "meta-regulation" has spread in such a relevant way that it became a *model*: for example, the evolution of the Basel I capital requirements to the Basel II, where instead of a prescriptive approach, simple and common to all banking institutions, a model of

---

[74] *See* Bamberger, Kenneth A, *"Technologies of compliance: Risk and regulation in a digital Age"*, in Tex. L. Rev., 88, 2009, p. 672 ff.

[75] *See* Black, Julia, *"Paradoxes and Failures: New Governance Techniques and the Financial Crisis*" in The Modern Law Review, Vol. 75, No. 6, 2012, p. 1045 ff.

[76] *See* Black, Julia, *"Paradoxes…", cit.*, p. 1046.

[77] On meta-regulation, *see* Coglianese, Cary and Mendelson, Evan, *"Meta-regulation and self-regulation"*, in The Oxford Handbook of Regulation, Oxford, Oxford University Press, 2010 (comparing traditional "command and control-based" regulation to "meta-regulation" and "self-regulation", whose non-consensual definitions are then presented), and Scott, Colin, *"Regulating everything: From mega- to meta-regulation"*, in Administration, Vol. 60, 2012, p. 57 ff.

[78] The expression belongs to Parker, Christine, *The Open Corporation: Effective self-regulation and Democracy*, Cambridge University Press, Cambridge, 2002, p. 245 ff., in which the author defends the so-called "open corporation", a company that "democratically self-regulates" in a fusion of management, democracy, and law. *See* also Parker, Christine, *"Meta-Regulation: Legal Accountability for Corporate Social Responsibility?"*, in The New Corporate Accountability: Corporate Social Responsibility and the Law, Cambridge University Press, Cambridge, 2007, p. 3.

adjustment was adopted on the basis of a process of interaction with the institution itself[79]. On a national level, one finds that the example of the legal framework built around the prevention of money laundering (*i.e.* Law 83/2017) unquestionably follows this model: each entity must effectively create and apply the policies, procedures and control mechanisms adequate to the capable management of risks related to money laundering which the company is or may find itself to be exposed to [Article 12, Paragraph 1, Section a)]. And it's the entity's own duty to identify, evaluate and mitigate such risks, for the purpose of which it must take into account its own specific characteristics (such as the size and complexity of its activity, its clients and their own activity, the countries or territories of origin, etc.) (*see* Article 14 of Law 83/2017). In its wake, several normative instruments issued by the Bank of Portugal, such as Notice 2/2018 – observe the vast array of rules therein appealing the entity to carry out an adequacy finding with regard to procedures, processes, means, etc. [*e.g.* Article 1, Paragraph 1, Sections c) and j); Article 7, Paragraph 1; Article 10, Paragraph 1; Article 15, Paragraph 2, Section c); and Article 19, Paragraph 2] – and Instruction 2/2021 [*e.g.* Article 5, Paragraph 3, Section c) and Article 17, Paragraph 1] rest on the same model by calling on the entity to set up the processes, procedures, and means adequate to reach the objectives laid down by the regulator.

In effect, insofar as it dictates that the regulated entity must lay down plans which adequately deal with its risk environment, this ("management-based") regulatory model has meant the increasing adoption of technology with the view of handling and creating the information necessary to model the risk in each organization and keep the processing of said information permanently updated.

Yet, *RegTech*'s large development within the span of the last decade is the result of specific reasons. First and foremost, it's a result of the *2007-2008 financial crisis*, which brought about a lot of regulatory demands that could be fulfilled (only) through the use of technology[80]. It's also a result of *financial regulation's own complexity*, which has

---

[79] *See* Chiu, Iris H-Y, *"Regulating…",* cit., p. 22 ff. (with several examples).

[80] Highlighting this reason in particular as the reason for *RegTech*'s development, *see* Arner, Douglas W., Barberis, Janos and Buckey, Ross P., *"FinTech..."*, cit., p. 395, and Arner, Douglas W., Barberis, Janos Nathan and Buckley, Ross P.,

meant increased demands on compliance[81]. Secondly, the great developments in the field of *data science*, namely the possibility of transferring computing to "cloud" infrastructures, also boosted *RegTech*. Thirdly, the pressure to *reduce costs* has equally meant opting for *RegTech* due to the savings it enables[82] – one ought to keep in mind that the estimated cost of AML compliance programs in the European Union already totaled 83 billion dollars in 2017[83]. All of this is taking place at a stage when banks are providing an increasingly digital experience, from which AI may emerge[84].

"The emergence of RegTech 2.0: From know your customer to know your data", cit., p. 9 ff.

[81] In this regard, *see* Lin, Tom C. W., *"Compliance..."*, cit., p. 166 ff., and Arner, Douglas W., Zetzsche, Dirk A., Buckley, Ross P. and Weber, Rolf H., *"The Future of Data-Driven Finance and RegTech: Lessons from EU Big Bang II"*, in Stanford Journal of Law, Business & Finance, Vol. 25, No. 2, 2020, p. 247.

[82] Kurum, Esman, *"RegTech solutions and AML compliance: what future for financial crime?"*, in Journal of Financial Crime, ahead-of-print, 2020, p. 3, identifies two reasons for the massive adoption of *RegTech*: not only cost reduction but also the long-term value it creates for institutions.

[83] *See* Kurum, Esman, *"RegTech solutions and AML compliance: what future for financial crime?"*, cit., p. 2. Other authors also state that, in the United States of America, the costs of fines imposed on financial institutions after the 2007-2008 financial crisis were over 200 billion dollars (*see* Arner, Douglas W., Barberis, Janos Nathan and Buckley, Ross P., *"The emergence of RegTech 2.0: From know your customer to know your data"*, cit., p. 2.); other sources say the cost went as high as 321 billion dollars in the years between 2008 and 2016 (43 billion dollars in 2016 alone) (*see* Fruth, Joshua, *Anti-money laundering controls failing to detect terrorists, cartels, and sanctioned states*, March, 2018, p. 3; *see also* Jung, John Ho Hee, *"RegTech..."*, cit., p. 258 ff., containing information with regard to the United Kingdom). In 2018, *Deloitte* estimated that the cost of compliance was 25 billion dollars in the United States of America alone (*see "The case for artificial intelligence in combating money laundering and terrorist financing. A deep dive into the application of machine learning technology"*, in Deloitte, 2018, p. 4) and *JP Morgan* spends about 600 million dollars a year on technology used for compliance (*see* Lin, Tom C. W., *"Compliance..."*, cit., p. 166). Today, the costs of "governance, risk, and compliance" ('GRC') represent 15% to 20% of the total costs of major financial institutions (*see* Jung, John Ho Hee, *"RegTech..."*, cit., p. 258). For a general sense of the costs associated with regulation, *see* Arner, Douglas W., Barberis, Janos and Buckey, Ross P., *"FinTech..."*, cit., p. 388 ff. And, most recently, the EBA's *Study of the Cost of Compliance with supervisory reporting requirements*, 2021 (Report EBA/Rep/2021/15).

[84] Noting this, *see* Armstrong, Patrick, *"Developments in RegTech and SupTech"*, in European Securities and Markets Authority, 2018, p. 2.

According to the data available, *RegTech* is in marked expansion. In the United Kingdom, for example, about 10 companies were started in that field in the year 2000; between 2010 and 2020, a minimum of 15 such companies were started in each year, with some years (such as 2016) seeing the start of almost 30 new companies. A steep decline in new companies has been seen recently, which may be attributed to the fact that the already existing ones are gaining a relevant size. The market is composed of an increasingly larger percentage of mature companies (more than 5 or even 10 years old)[85]. *FinLab*, the platform created by the Portuguese financial supervisors (the Bank of Portugal, the Securities Exchange Market Commission, and the Supervising Authority for Insurance and Pension Funds) identified 16% of projects in the field of *RegTech* in its report of the second edition of *Portugal FinLab*, in 2020 (in its first edition, in 2019, it had identified 13% of projects)[86].

The areas served by *RegTech* are mostly concentrated around matters of compliance: 32% of products regard financial crimes (AML) – for instance, HSBC recently announced an agreement with "Silent Eight" for the development of AI mechanisms; *Standard Chartered* announced a similar agreement with "Quantexa"[87]; 16.5% regard data protection and privacy; and 9% regard management and regulatory compliance[88]. According to other sources, over half of all *RegTech* companies in 2017 focused on AML compliance. In the *RegTech 3.0* era, it's expected that the focus will be on the increasing importance of data for AML compliance ('know your data')[89].

---

[85] All these elements may be found in *"2021: A Critical..."*, cit., p. 17.

[86] *See Portugal Finlab Report*, 2nd Edition, 2020, p. 8 (available at https://8080dd92-d6fc-49d9-a97eb24c8f013bb2.filesusr.com/ugd/ca9a53_217c4187d5bd4a5a9b377c6f6500e0ff.pdf).

[87] *See "2021: A Critical..."*, cit., p. 16

[88] *See "2021: A Critical..."*, cit., p. 13 ff., and *"There's a revolution coming. Embracing the challenge of RegTech 3.0"*, in KPMG, 2018, p. 1 ff.

[89] In this regard, *see* Kurum, Esman, *"RegTech solutions and AML compliance: what future for financial crime?"*, cit., p. 2. A description of the areas where *RegTech* most intervenes and of the technologies it most uses [such as AI, machine learning, robotic process automation ('RAP'), natural language processing ('NPL'), big data, cloud computing, etc.] may be read in Jung, John Ho Hee, *"RegTech..."*, cit., p. 265 ff., and also in the important report *"Machine learning in UK financial services"*, in Bank of England, 2019.

IV. TECHNOLOGY (*INTER ALIA*, AI) IN BANKING COMPLIANCE

Unsurprisingly, the financial sector, which has always been an avid user of technical and technological innovations[90], is at the forefront of developing uses for them. And the advantages that the sector may reap by using AI are clear[91]. The fact that AI is particularly suitable to be used by the financial sector explains the significant attention recently paid by national and international entities, by regulators, etc., to this matter in specific[92].

Compliance is commonly named as one of the areas of banking activity *most suitable to the use of AI* – what's more, compliance has always had a close bond with technology due to it being a "back office"

---

[90] In this regard, *see* Maia, Pedro, *"A robotização do mundo financeiro: reflexões introdutórias"*, in Estudos de Direito do Consumidor, No. 16, Centro de Direito do Consumo - Instituto Jurídico, Coimbra, 2020, p. 273 ff.

[91] *See, for example, "EBF position paper on AI in the banking industry"*, in European Banking Federation, 2019, EBA, *Report on big data and advanced analytics*, 2020, p. 43 ff., EBA, *EBA Analysis of Regtech in the EU Financial Sector*, 2021.

[92] As an example, *see* the EBA *Report on automation in financial advice*, 2016, (available at https://esasjointcommittee.europa.eu/Publications/Reports/EBA%20BS%202016%20422%20(JC%20SC%20CPFI%20Final%20Report%20on%20automated%20advice%20tools).pdf), the EBA *Report on big data and advanced analytics*, 2020 (available at https://www.eba.europa.eu/sites/default/documents/files/document_library//Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf), the ESMA – *Joint Committee Final Report on Big Data*, 2018 (available at https://www.esma.europa.eu/sites/default/files/library/jc-2018-04_joint_committee_final_report_on_big_data.pdf), the *EBF position paper on AI in the banking industry*, 2019, (available at https://www.ebf.eu/wp-content/uploads/2020/03/EBF_037419-Artificial-Intelligence-in-the-banking-sector-EBF.pdf), the *Machine learning in UK services*, 2019, issued by the *Bank of England* and the FCA (available at https://www.bankofengland.co.uk/report/2019/machinelearning-in-uk-financial-services), Calzolari, G., *Artificial Intelligence market and capital flows, Study for the Special Committee on Artificial Intelligence in a Digital Age*, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2021 (available at https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662912/IPOL_STU(2021)662912_EN.pdf). For an approach that's not purely sectorial, *see Livro Branco da Comissão Europeia sobre a inteligência artificial – Uma abordagem europeia virada para a excelência e a confiança*, 2020 (available at https://op.europa.eu/pt/publication-detail/-/publication/ac957f13-53c6-11eaaece-01aa75ed71a1), the recent proposal of the European Commission for an *Artificial Intelligence Act* (available at https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0004.02/DOC_1&format=PDF).

function[93]; this is first and foremost true of AML compliance, but also of regulatory compliance[94]. Entities are gradually subject to more KYC ('know your customer') duties, whose efficiency can be greatly increased if the information on which they rest – to know a client is to know *information* about the client – can be cross-checked and cross--referenced between different sources (beginning with the information provided by the client themselves) on a large scale in a short period of time, or even almost instantaneously. Moreover, the paradigm has shifted as AML compliance's methodology ceased to rest on the *client* and turned to *data* – "data is king"[95]. Since credit institutions possess an (exponentially) increasing volume of data which they have the burden of adequately using – starting with assessing its quality and authenticity[96] –, technological solutions have become crucial and of growing usefulness. In this regard, application programming interfaces are able

---

[93] *See* Fanto, James A., *"The Professionalization of Compliance: Its Progress, Impediments, and Outcomes"*, in Notre Dame Journal of Law, Ethics & Public Policy, Vol. 35, No. 1, 2021, p. 223.

[94] *See*, for example, Magnuson, William, *"Artificial Financial Intelligence"*, in Harvard Business Law Review, Vol. 10, 2020, p. 350, Kaya, Orçun, *"Artificial intelligence in banking: A lever for profitability with limited implementation to date"*, in Deutsche Bank Research, 2019, p. 5, and *see*, most recently, the empirical data in *"2021: A Critical..."*, cit., p. 13 ff., as well as EBA, *EBA Analysis...* cit., p. 42 ff.

[95] *See* Arner, Douglas W., Barberis, Janos Nathan and Buckley, Ross P., *"The emergence of RegTech 2.0: From know your customer to know your data"*, cit., p. 16 ff., and Kurum, Esman, *"RegTech solutions and AML compliance: what future for financial crime?"*, cit., p. 3. Besides, from a very interesting perspective, computers have something in common with cells and with the human brain: in different ways, all are processors of *information* (*see* the inspiring work of Oliveira, Arlindo, *The Digital Mind: How Science is Redefining Humanity*, MIT Press, Cambridge, 2017, p. 1).

[96] It's important to keep in mind that, under the terms of the Bank of Portugal's Notice 2/2018, institutions have the *duty* of resorting to various sources of information (Article 6), first and foremost internal ["Analysis and internal documents of financial entities, including information collected during the procedures of identification and diligence and the lists and databases internally produced and updated – Paragraph 2, Section g)], but also external, where "Independent and credible information from civil society or international organizations [Paragraph 2, Section h)] is included, and "Information gathered from the internet and mass media, as long as belonging to a credible and independent source" [Paragraph 2, Section i)], the information contained in databases, lists, risk reports, and other analysis originating in commercial sources available in the market [Paragraph 2, Section j)], official statistical data from national or international sources [Paragraph 2, Section k)].

to produce great results not only at onboarding [Article 23, Paragraph 1, Section a) of Law 83/2017] but also with regard to the permanent *update* of the information which bounds entities (*see* Article 40 of Law 83/2017).

Likewise, credit institutions have in *RegTech* a valuable ally in defining and updating each client's *risk profile* [Article 18, Paragraph 2, Section c) of Law 83/2017] based on the information collected.

Still in connection with the prevention of money laundering, pursuant to Article 39 of Law 83/2017 credit institutions hold duties in respect of "politically exposed persons"[97]: they're charged with identifying a politically exposed person [Article 39, Paragraph 1, Section a) of Law 83/2017] and thereafter subject that person's operations to the very strict applicable law and jurisdiction. Further duties regard "entities to which sanctions have been applied", whose funds and economic resources have been subject to restrictive measures by the United Nations or European Union (*see* Article 13 and following of Law 97/2017). The challenges posed to banks are truly massive[98] due to the necessity to screen the names of the transacting parties and to cross-check those with the ones included on the lists: contrary to the names used on the lists (of politically exposed persons or of persons to whom sanctions have been applied), in transactions names may appear as abbreviations, initials, with or without full last names (or even in reverse order) – one must not forget these are worldwide lists, made up of persons of all nationalities and languages –, together with the use of homonyms (the more incomplete the name used in the transaction is, the greater the use will be) , which all in all makes AI's ability to make the screening more flexible all the more useful. A "rule-based" system is either too strict – and will no longer detect the entity should even the slightest difference exist in its identification – or too comprehensive, in which case it will generate an inordinate amount of false positives.

Given the constant change of the universe of persons qualifiable as politically exposed and to whom sanctions have been applied, and the fact that their number is vast to begin with, it's easy to understand

---

[97] Defined by a (decisive) list, in Article 2, Paragraph 1, Section cc) of Law 83/2017.

[98] For an international reference to the challenges and costs of implementing these regimes, *see* Arner, Douglas W., Barberis, Janos and Buckey, Ross P., *"FinTech..."*, cit., p. 391.

RegTech's usefulness in ensuring compliance. As a matter of fact, given that a person's qualification as politically exposed determines which legal framework will be applied to the transaction itself, it can be said that credit institutions would find it difficult to screen operations within a reasonable timeframe should they simply have do it by hand.

Furthermore, again in connection with the prevention of money laundering *RegTech* has increasingly (and even decisively) assisted in fulfilling the duty to analyze, exam [Article 11, Paragraph 1, Section g) and Article 52 of Law 83/2017] and report suspicious operations [Article 11, Paragraph 1, Section c) and Article 43 and following of Law 83/2017] by allowing entities, mostly through the use of AI (of a subset of AI in particular: *machine learning*[99]*), to identify their clients' suspicious activities*[100] *and, making use of ample databases, anomalies as well. AI is almost unavoidable in precluding the difficulties associated with automated systems (whose assessments and warnings are the result of a closed set of rules): they generate a huge number of "false positives"*[101]*, leaving a rather significant number of operations to be assessed by human persons*[102].

The (growing) use of technology is not only partially spontaneous, a result of the credit institution's need to meet its operative interests, but also to a great extent the regulator's de facto imposition, in the sense that it imposes on the institution demands which can only be met with the use of AI. A persuasive example of this was seen when the German regulator demanded that, in a relatively short period of time, a credit institution reassess 20 million financial operations it had made in the

---

[99] An explanation of machine learning may be found in Domingos, Pedro, *The master algorithm: How the quest for the ultimate learning machine will remake our world*, Basic Books, 2015, p. 5 ff. ("Every algorithm has an input and an output: the data goes into the computer, the algorithm does what it will with it, and out comes the result. Machine learning turns this around: in goes the data and the desired result and out comes the algorithm that turns one into the other. Learning algorithms — also known as learners — are algorithms that make other algorithms"). Given its great development and importance for AI, there is a tendency to associate one with the other, although this assimilation is incorrect. *See* Scopino, Gregory, *"Key...",* cit., p. 23.

[100] *See*, for example, *"The case..."*, cit., p. 9.

[101] *See* Fruth, Joshua, *Anti-money laundering controls failing to detect terrorists, cartels, and sanctioned states*, cit., p. 3.

[102] *See* Kurum, Esman, *"RegTech solutions and AML compliance: what future for financial crime?"*, cit., p. 5.

past from a money laundering perspective[103]. It would only have been feasible to comply with the order by creating an AI-run tool, which is what ended up being done.

Besides, the goal of inducing credit institutions to use AI for the purpose of AML compliance is freely acknowledged in the field of regulation[104], without prejudice to the *principle of technological neutrality*. Such principle may take three different directions: (i) it can mean that the technical requisites for avoiding negative externalities (such as pollution, radio interference, etc.) are designed by defining the end result, all the while granting companies the freedom to choose the technology most appropriate to reach it; (ii) it can mean that those same regulatory principles are applicable regardless of the technology used by the regulated entity; or (iii) it can mean that regulators themselves should avoid using regulation as a means of steering the market to a certain structure which they deem optimal[105]. When taking into account the economic implications of the intensive use of technology – due to the scale economies it enables –, regulatory demands imposing the use of such technologies may surely lead to changes in the market's structure. It is therefore possible that the principle of technological neutrality will be reviewed and made flexible in a way that limits it to neutrality with regard to the "seller of technology" but not with regard to any other aspects[106]. *RegTech*'s advances in regulatory compliance and the increased use of technology articulated between the regulators and the regulated entities may in future require a certain harmonization of technological solutions, which will somewhat limit the principle of technological neutrality.

The benefits linked to AI and its associated technologies are many: AI offers the possibility of analyzing, screening, etc., the *complete*

---

[103] *See* Zimiles, Ellen, "How AI is transforming the fight against money laundering", *World Economic Forum*, 2019 (available at https://www.weforum.org/agenda/2019/01/how-ai-can-knock-thestarch-out-of-money-laundering).

[104] A report regarding the position of various regulators of favoring or stimulating the use of AI in AML compliance can be read in Estrada, Juan Carlos, *"The AML Arms Race: How Artificial Intelligence and Machine Learning Will Combat Money Laundering"*, in Rutgers Bus. LJ, 16, 2020, p. 393 ff.

[105] *See* Maxwell, Winston J and Bourreau, Marc, *"Technology neutrality in internet, telecoms and data protection regulation"*, in Computer and Telecommunications Law Review, 31, 2014, p. 1.

[106] *See "2021: A Critical..."*, cit., p. 19 ff.

universe of operations – regardless of their amount, the place where they're ordered, the jurisdiction to which their beneficiaries belong, the time and day of the week when they take place, etc. – *in real time* – for example, by blocking a credit card payment operation – through a collection of data ('big data') inaccessible to human knowledge. It is not just a benefit; it's also an *inevitability* if AML compliance is to be in any way effective in the face of the current financial situation: fully global, facing a growing use of electronic payments, and with an exponentially increased flow of goods. One should bear in mind the occasion when sales on *eBay*, paid for with *PayPal* and used to launder money for the Islamic State, went undetected[107]. It's simply not possible to fulfill the objective of AML compliance using human resources only. It may therefore be said that the development of technology both fuels money laundering and offers a solution to the problem[108].

V. RISKS AND CHALLENGES OF AI IN BANKING COMPLIANCE

Now that it has been established that the use of AI in AML (and regulatory) compliance tends to be inevitable[109], the risks[110] associated

---

[107] The evolution of the methods used by criminal networks for money laundering (namely, to upload it to the financial system) is huge and poses immense challenges to both the financial sector and compliance systems. For a description of these methods, *see* Miller, Geoffrey P., *"The Role of Risk Management and Compliance in Banking Integration"*, in NYU Law and Economics Research Paper, 14-34, 2014, p. 44 ff.

[108] *See* Estrada, Juan Carlos, *"The AML Arms Race: How Artificial Intelligence and Machine Learning Will Combat Money Laundering"*, cit., p. 386.

[109] Expressly in this sense, *see* Estrada, Juan Carlos, *"The AML Arms Race: How Artificial Intelligence and Machine Learning Will Combat Money Laundering"*, cit., p. 400, and, in respect of the use of *RegTech* for compliance, *see* Kurum, Esman, *"RegTech solutions and AML compliance: what future for financial crime?"*, cit., p. 3.

[110] The *risks* and, in truth, the *limitations* as well: as writes Packin, Nizan Geslevich, "*RegTech, Compliance and Technology Judgment Rule"*, in Chicago-Kent Law Review, Vol. 93, No. 1, 2018, p. 194, *RegTech* is not a cure-all for every problem. Artificial Intelligence systems used for compliance may succeed in identifying and reporting (regulatory or money laundering) breaches, but are very limited in creating a culture of compliance. And they may even become what the author calls *"anti-regtech" – the manipulation of technology to forge compliance with regulatory demands*.

with such use must be outlined. These risks have different natures and are at different levels.

Firstly, there exists the risk of the algorithm malfunctioning[111] as a result of a flawed or incorrect design. It's true that with algorithms, as with any other good or service, an error may occur. But here two significant particularities greatly aggravate the risk of that happening.

Secondly, the effects of an algorithm's imperfection tend to be exponentially aggravated: unlike human error[112], which is inclined to be limited to a (minority) share of each person's actions and is therefore *individual* and *partial*, an algorithm's error is inclined to be *universal* and *whole* since it will affect all of its activity and not just one part of it. If the same algorithm is already prevalent in the market and is used by several credit institutions, one sole mistake can have systemic repercussions. Technology's deficiencies or compromises may thus have universal consequences[113].

Thirdly, and of equal importance, detecting an error may be much harder – in some cases, it may even be impossible. Since AI feeds off big data, whose true extent is inaccessible to human knowledge, it becomes very difficult to recognize that, based on the information available ("unknown" to human persons on account of its magnitude), the algorithm has made wrong or inappropriate decisions.

This is one of the chief risks of AI: the data used to make decisions. The issues are many: the data might be incomplete because it was collected from a limited universe of samples, in which case the algorithm will be compromised due to the fact that, for example, it will draw conclusions about a certain universe from a distinct or far

---

See Packin, Nizan Geslevich, *"RegTech..."*, cit., p. 212 ff. On the risks of AI in the financial sector, *see*, most recently, EBA, *EBA Analysis...,* cit., p. 38 ff.

111 In layman's terms, "an algorithm is a sequence of instructions telling a computer what to do" (*see* Domingos, Pedro, *The master algorithm: How the quest for the ultimate learning machine will remake our world*, cit., p. 1).

112 The risk of AI elevating human errors may also be identified. In this sense, *see* Magnuson, William, *"Artificial..."*, cit., 125), p. 340 ff. ("the greatest danger of artificial intelligence is not that of exceeding human intelligence, but of exacerbating human error").

113 *See* Bamberger, Kenneth A, *"Technologies of compliance: Risk and regulation in a digital Age"*, cit., p. 710 ff. (highlighting that the effects of "codifying" the algorithm are much like those of the law itself, which generalizes its applicability, creating a framework which persists over a long time).

away universe of samples; the data might contain *mistakes*[114] *(which of course harms the quality of AI's output: "garbage in, garbage out"*[115]*); the data might be (partially) false*, whether it be because fake news have been spreading on social media or because hackers have "poisoned data" so as to influence the AI's judgement – problems which can only be overcome by way of cleansing processes, exceedingly expensive because of the need to use massive human resources and therefore with a tendency to be avoided[116]; data might be *outdated*, in the sense that it does not correspond to the current reality; data might be a "compromised piece" of reality conveying the views or perceptions of society, or of a part of society – if the data includes news reports (and for the purpose of AML compliance it usually does) it's important to consider that mass media follows editorial guidelines, that journalists choose what to report, etc. A very telling example is that of the algorithm which, while using big data to recruit an employee, presumed that the employer preferred to hire men over women and thus proceed to reject every female candidate to the job. It all depends on data and on the conclusions – the patters and models – drawn from it by the algorithm[117]. The risk that AI may create instances of discrimination have been highlighted by theorists[118], with some authors going as far as saying that this side-effect is intrinsic to the prediction itself.

---

[114] The number of errors in reports from technical sources is surprising: in 2004, the *National Association of State Public Interest Research Groups* assessed that 79% of reports contained mistakes, 25% contained serious mistakes, 54% contained imprecise personal information, and 30% listed closed accounts as still active. *See "The case..."*, cit., p. 522 ff.

[115] *See* Buckley, Ross P., Zetzsche, Dirk A., Arner, Douglas W. and Tang, Brian W., *"Regulating..."*, cit., p. 50.

[116] *See* Buckley, Ross P., Zetzsche, Dirk A., Arner, Douglas W. and Tang, Brian W., *"Regulating..."*, cit., p. 51.

[117] Data related to *Enron* – the company which was at the center of one of the biggest and most serious scandals of accounting fraud and information forging – was used to feed compliance algorithms. *See* Enriques, Luca and Zetzsche, Dirk A., *"Corporate Technologies and the Tech Nirvana Fallacy"*, in European Corporate Governance Institute (ECGI), No. 457, 2019, p. 25.

[118] *See*, with updated information, Magnuson, William, *"A Unified..."*, cit., p. 25 ff.

It's important to note that, when handling "big data", AI doesn't use information *as it is*, in its context[119]: algorithms necessarily disaggregate information into "pieces" only to re-aggregate it immediately afterwards and establish links between features which are in no way interconnected in real life. For example, absurd though it might sound, if an analysis of data demonstrates that more suspected money laundering operations take place between eight-thirty and nine-thirty in the morning, the algorithm will establish a link between time and money laundering and will start to consider the time when the operation takes place as an assessment criteria. Many more examples (even stranger and more absurd) may be thought of. Strictly speaking, the information dealt with by AI isn't *existing information*; it's *constructed information*, in the sense that associations which do not *actually* exist are created and established – associations which amount to an intellectualization of reality. What's more, information, in its full dimension and completeness, is something which exists *only for the machine*; it does not exist for human persons because they are incapable of knowing, processing and associating it with the vastness of data that, aided by supercomputers, the algorithm takes into account when making decisions.

In addition, although "big data" is information – which in and of itself doesn't represent anything new or specific – its characteristics greatly differ from those of common (traditional) information, something which makes them qualitatively different and poses specific problems: their *magnitude* – there is more data than ever before and it's being produced at an unprecedented rhythm; their *permanence* – data persists in time and may be stored indefinitely; and their *portability* – data may be copied, transferred, shared, and stolen[120].

One must not presume that the existence of a great magnitude and quantity of information means it's freely accessible. The fact that accessing to (constructed, aggregated, etc.) information tends to come at

---

[119] It must be highlighted that the data used by AI is not limited to existing or available data; data may be created for this purpose. For example, when a start-up company employs about 30 thousand workers to catalogue real-life images and then sells the data thus created to be used by artificial intelligence systems such as self-driving. *See* Magnuson, William, *"A Unified..."*, cit., p. 32.

[120] *See* Magnuson, William, *"A Unified..."*, cit., p. 29 ff.

an expensive price[121] in and of itself raises questions, especially when the access might be relevant to public interest, as is the case with AML compliance.

Besides risks related to *data*, there are (many) more related to the *algorithm* itself. AI is capable of learning *supervised* or *unsupervised*. Learning is *supervised* when the algorithm learns from a previously catalogued collection of data: for example, when the operations recorded in the database which the algorithm used as a starting point had already been classified as suspicious or not[122]. In a system such as this, the quality of the information (of the classification) is essential: if the information used for learning is incorrect or incomplete the algorithm may ultimately draw wrong conclusions[123].

On the other hand, learning is *unsupervised* when it rests on free data and takes place without previous training[124]. Although the risks associated with this method are clear, it ought not to be rejected on account of that because supervised learning will, in principle, prevent the algorithm from learning and identifying standards *different* from those underlining the collection of data used for training. Returning to the example used above, if criminal networks resort to a new method of money laundering – and they are always seeking to devise new ways unknown to authorities – that means the algorithm which

---

[121] *See* Buckley, Ross P., Zetzsche, Dirk A., Arner, Douglas W. and Tang, Brian W., *"Regulating..."*, cit., p. 49 ff.

[122] A common example found in literature, in the instance where it's intended that the algorithm identifies the image of a cat, consists of creating a database of images classified as "cat" or "non-cat" so that the algorithm then classifies other images. The programmer does not indicate the meaning of cat, or the determining elements of a cat's image; he or she simply ensures that the images used for "learning" have been correctly classified as "cat" or "non-cat". *See* Scopino, Gregory, *"Key..."*, cit., p. 30 ff.

[123] *See* Scopino, Gregory, *"Key..."*, cit., p. 32. The following example is given: if, in the collection of data made available to the system, all words ending with "ing" are classified as verbs – because the collection neither contain nouns (such as "king") nor adjectives (such as "interesting") ending with "ing" –, then the system will classify all words ending with "ing" as verbs in the future.

[124] In this regard, *see*, for example, Johnson, Kristin, Pasquale, Frank and Chapman, Jennifer, *"Artificial intelligence, machine learning, and bias in finance: toward responsible innovation"*, in Fordham L. Rev., 88, 2019, p. 506 ff., and Scopino, Gregory, *"Key..."*, cit., p. 30 ff. (who further differentiates "reinforcement learning").

learned under supervision will not (or will hardly be able to) identify an operation as suspicious, since an operation of that kind and the corresponding standard of suspicion were not present in the database which it was provided.

In addition, AI is rather *complex* and *opaque*, with its working model being called a "black box"[125] which poses the serious risk of resting on processes and operations unknown to persons (or even inaccessible to human knowledge) and therefore out of their respective control[126]. In truth, algorithms go through a huge collection of data, identify certain relationships or patterns, generate new standards with which to assess new data, etc. This working model makes it very difficult or even impossible to tangibly reconstruct the process leading up to the algorithm's decision[127]: this is what's called AI's *unpredictability*, also known as *unknowability* or *cognitive unaccountability*[128]. *Two risks arise therefrom: on the one hand, the inability of absolutely predicting* or anticipating the algorithm's future behaviors – there are no 100% safe algorithms[129]. On the other hand, the risk inherent to the (eventual) inability to demonstrate the reasoning behind the algorithm's decision creates many problems. First and foremost, it creates the problem of controlling the quality of its performance. Secondly, it creates a

---

[125] *See "The case..."*, cit., p. 507 ("an effort to explain [AI's] «reasoning» would be about as useful as a map of all the synapses and other chemical reactions in the brain that occur when, say, a manager decides whether to grant or deny an employee's request for a vacation day").

[126] his leads some authors to claim the need of including humans in the circuit of artificial intelligence. *See* Buckley, Ross P., Zetzsche, Dirk A., Arner, Douglas W. and Tang, Brian W., *"Regulating..."*, cit., p. 44 ff.

[127] *See*, for example, Estrada, Juan Carlos, *"The AML Arms Race: How Artificial Intelligence and Machine Learning Will Combat Money Laundering"*, cit., p. 401 ff. Setting "rule-based" AI – which rests on pre-determined rules and therefore allows for the explanation of decisions – against "machine learning" AI – which doesn't allow for the explanation of the reasoning behind its decisions, which are taken based on the identification of statistical correlations among the data –, *see* Kingston, John, *"Using artificial intelligence to support compliance with the general data protection regulation"*, in Artificial Intelligence and Law, Vol. 25, No. 4, 2017, p. 431 ff.

[128] *See* Yampolskiy, Roman V., *"Unpredictability of AI"*, in Cornell University, 2019, p. 2 (highlighting that the concept of artificial intelligence's *unpredictability* is related to, but not to be confused with, *unexplainability or incomprehensibility).*

[129] In this sense, *see* Yampolskiy, Roman V., *"Unpredictability..."*, cit., p. 5.

regulatory problem: banks must be able to prove that they comply with regulatory demands. If it's not possible for them to demonstrate the reasoning behind the algorithm's decision that ability is compromised. For that reason, a previous commitment by the regulators to consider the use of AI as the fulfillment of certain regulatory demands has been needed in some cases. What some authors call legal risk (or "translation problem"[130]) is different from this: regulations aren't "machine-readable", meaning that they must always be translated into the algorithm, at the risk of the regulator's deficient – or discrepant – interpretation and the subsequent contamination of all compliance activity with an interpretation against the regulatory framework[131]. And problems pertaining to the General Data Protection Regulation are plentiful as well, namely the data subject's right "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her" (Article 22), or the right to be forgotten[132].

AI is but a piece of a *system* ("AI's ecosystem")[133], which has been experiencing an unparalleled technological development and which is already is widely used in the financial sector. But this new technology is so disruptive that it will entail new approaches in regulation – it's demanded that the regulators themselves take up AI when performing their duties – and in controlling the competitive effects that said new technology might generate. And it will also entail the creation of a

---

[130] The expression belongs to Bamberger, Kenneth A, *"Technologies of compliance: Risk and regulation in a digital age"*, cit., p. 706.

[131] In this regard, *see* Chiu, Iris H-Y and Lim, Ernest Wk, *"Managing Corporations' Risk in Adopting Artificial Intelligence: A Corporate Responsibility Paradigm"*, in Wash. U. Global Stud. L. Rev., 20, 2021, p. 366 ff.

[132] On questions raised by the general framework of data protection, *see* Buckley, Ross P., Zetzsche, Dirk A., Arner, Douglas W. and Tang, Brian W., *"Regulating..."*, cit., p. 58 ff., Arner, Douglas W., Zetzsche, Dirk A., Buckley, Ross P. and Weber, Rolf H., *"The Future..."*, cit., p. 256 ff., Kingston, John, *"Using..."*, cit., p. 439 ff., Kaya, Orçun, *"Artificial..."*, cit., p. 6, Chiu, Iris Hy and Lim, Ernest Wk, *"Managing Corporations' Risk in Adopting Artificial Intelligence: A Corporate Responsibility Paradigm"*, cit., p. 367, and Lee, Joseph, *"Access to Finance for Artificial Intelligence Regulation in the Financial Services Industry"*, in European Business Organization Law Review, Vol. 21, 2020, p. 745.

[133] *See* Giuffrida, Iria, *"Liability for AI Decision-Making: Some Legal and Ethical Considerations"*, in Fordham Law Review, Vol. 88, 2019, p. 442.

legal framework which regulates the use of AI while both assuming its inevitability and the need to safeguard certain essential values[134] (already under preparation, as evidenced by the proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence, *i.e.* the Artificial Intelligence Act).

A final reference should be made with regard to the risks of the (*RegTech*) market on which the provision of AI-related services rests. As previously stated, these can threaten the financial system. There's a strong tendency towards concentration in these markets, whether because financial institutions give preference to larger or more "mature" service providers – so as to reduce the risk posed by the service's shutdown[135] – or due to the existence of scale economies on the side of the provider: in 2018, the four largest cloud service providers held an 80% share of the world market and 25% of banks' core systems were stored in clouds[136]. There's a fear that the collapse of only one of these service providers may cause a worldwide disruption of banking systems. Usually, the extraordinary complexity of AI services and the outsized cost of developing them mean that they tend to be outsourced by institutions, which increases the risk of dependence on third parties, all the greater due to the market's concentration. Even though several steps were taken towards minimizing the systemic risks of financial institutions after the 2007-2008 financial crisis, it may be that a larger and unmitigated risk is developing with regard to the outsourcing of AI systems, clouding, etc. These risks are also intimately linked to intelligent compliance and need to be mitigated.

The evolution of traditional banking into "data-driven finance"[137] entails structural changes to the operation of banks, to the risks they are exposed to, and (maybe above all) to the risks which they expose

---

[134] The need to establish a framework of principles which artificial intelligence must abide by has been welcomed and stated by many. For an updated report and discussion on the possible or already implemented frameworks, *see* Buckley, Ross P., Zetzsche, Dirk A., Arner, Douglas W. and Tang, Brian W., *"Regulating..."*, cit., p. 57 ff., and Solow-Niederman, Alicia, *"Administering Artificial Intelligence"*, in S. Cal. L. Rev., 93, 2019, p. 635 ff.

[135] Identifying this fact and its implications, *see "2021: A Critical..."*, cit., p. 41.

[136] *See* Jung, John Ho Hee, *"RegTech..."*, cit., p. 269.

[137] The expression has been used by Arner, Douglas W., Zetzsche, Dirk A., Buckley, Ross P. and Weber, Rolf H., *"The Future..."*, cit., p. 245 ff.

third parties to, namely clients and citizens. This new stage is characterized by a strong interdependence between operation and frameworks: data protection frameworks, open banking frameworks, digital identification frameworks, and regulatory frameworks. Considering the European Union's all-encompassing interventions in 2018, it's not out of line to talk of a "Big Bang" in *RegTech* and "data-driven finance"[138].

In spite of having technology – namely AI – progressively more at its service, compliance will not go without the persistence of human intervention when it comes to two key aspects: the interpretation of regulatory frameworks – which withstand the "encoding" of algorithmic systems –, the observance of a culture of compliance within the organization, and the interpretation of compliance's development needs[139]. An (urgent) awareness of the risks associated with the massive introduction of AI (machine learning, in particular) in banking is necessary. Because of its complexity, speed, opaqueness, and interconnection, AI exposes the financial system to new and significant risks and thus makes it even more fragile, a "driverless" financial system with all associated risks[140].

REFERENCES

Allen, Hilary J., "Driverless Finance", in *Harvard Business Law Review*, Vol. 10, 2020, 158-206

Anagnostopoulos, Ioannis, "Fintech and regtech: Impact on regulators and banks", *Journal of Economics and Business*, Vol. 100, 2018, 7-25

Armour, John / Garrett, Brandon L./ Gordon, Jeffrey N. / Min, Geeyoung, "Board Compliance", *Minnesota Law Review*, Vol. 104, 2019, 1191-1273

---

138 *See* Arner, Douglas W., Zetzsche, Dirk A., Buckley, Ross P. and Weber, Rolf H., *"The Future..."*, cit., p. 247 ff.

139 *See* Chiu, Iris H.-Y. and Lim, Ernest W. K., *"Technology vs Ideology: How Far will Artificial Intelligence and Distributed Ledger Technology Transform Corporate Governance and Business?"*, in Berkeley Business Law Journal, Vol. 18, No. 1, 2021, p. 14.

140 *See* Allen, Hilary J., *"Driverless Finance"*, in Harvard Business Law Review, Vol. 10, 2020, p. 158 ff.

Armour, John / Gordon, Jeffrey / Min, Geeyoung, "Taking Compliance Seriously", in *Yale Journal on Regulation*, Vol. 37, No. 1, 2020, 1-66.

Armstrong, Patrick, "Developments in RegTech and SupTech", in European Securities and Markets Authority, 2018, 1-7.

Arner / Barberis / Buckley, "The emergence of RegTech 2.0: From know your customer to know your data", *Journal of Financial Transformation*, vol. 44, 2016, 79-86

Arner, Douglas W., / Barberis, Jànos / Buckley, Ross P., "FinTech, RegTech, and the Reconceptualization of Financial Regulation", *Northwestern Journal of International Law & Business*, Vol. 37, No. 3, 2016, 371-414.

Arner, Douglas W. / Barberis, Jànos / Buckley, Ross P., "Fintech and Regtech in a Nutshell, and the Future in a Sandbox", *CFA Institute Research Foundation*, 2017, 1-20

Arner, Douglas W./ Buckley, Ross P./ Barberis, Janos N., "The Evolution of Fintech: A New Post-Crisis Paradigm?", *Georgetown Journal of International Affairs*, vol. 47, 2016, 1271-1319.

Arner, Douglas W./Zetzsche, Dirk A./Buckley, Ross P./Weber, Rolf H., "The Future of Data-Driven Finance and RegTech: Lessons from EU Big Bang II", in *Standford Journal of Law, Business & Finance*, vol. 25, n.º 2, 2020, 245-288.

Baer, Miriam Hechler, "Governing Corporate Compliance", *Boston College Law Review*, vol. 50, 2009, 949-1019.

Bainbridge, Stephen M., "Caremark and Enterprise Risk Management", in *The Journal of Corporation Law*, vol. 34, 2008.

Bamberger, Kenneth A, "Technologies of compliance: Risk and regulation in a digital age", in *Tex. L. Rev.*, 88, 2009, 669.

Bank Of England, "Machine learning in UK financial services", 2019.

Banner, Stuart, "What causes new securities regulation? 300 years of evidence", *Washington University Law Quarterly*, 75, N.º 2, 1997, 1-6.

Bastos, Nuno Moraes, "Corporate Governance, Compliance e a Função Compliance nos Setores Bancários e Segurador", in *A Emergência e o Futuro do Corporate Governance em Portugal*, vol. II, Almedina, Coimbra, 2018, 207-234

Baumanns, Charlotte, „Fintech als Anlageberater? Die aufsichtsrechtliche Einordnung von Robo-Advisory", BKR, 2016, 366-375.

Bebchuk, Lucian A./Tallarita, Roberto, "The Illusory Promise of Stakeholder Governance", in Paper SSRN, 2020, 1-68.

Black, Julia, "Paradoxes and Failures: New Governance Techniques and the Financial Crisis", *The Modern Law Review*, vol. 75, n.º 6, 2012, 1037.

Bradley, Christopher G., "Fintech's Double Edges", *Chicago-Kent Law Review*, vol. 93, n.º 1, 2018, 61-95.

Brummer, Chris/Yadav, Yesha, "Fintech and the Innovation Trilemma", *The Georgetown Law Journal*, vol. 107, 2019, 235-307.

Buckley, Ross P./Zetzsche, Dirk A./Arner, Douglas W./Tang, Brian W., "Regulating Artificial Intelligence in Finance: Putting the Human in the Loop", *Sydney Law Review*, vol. 43, n.º 1, 2021, 43-8.

Calzolari, G., "Artificial Intelligence market and capital flows, Study for the Special Committee on Artificial Intelligence in a Digital Age", Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2021 (available at https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662912/IPOL_STU(2021)662912_EN.pdf)

Chiu, Iris H.-Y. and Lim, Ernest Wk, "Managing Corporations' Risk in Adopting Artificial Intelligence: A Corporate Responsibility Paradigm", *Wash. U. Global Stud. L. Rev.*, 20, 2021, 349-389.

Chiu, Iris H.-Y., "Regulating (From) the Inside. The Legal Framework for Internal Control in Banks and Financial Institutions", Hart Publishing, Oxford, 2015.

Chiu, Iris H.-Y./Lim, Ernest W. K., "Technology vs Ideology: How Far will Artificial Intelligence and Distributed Ledger Technology Transform Corporate Governance and Business?", *Berkeley Business Law Journal*, vol. 18, n.º 1, 2021, 1-63.

Coffee, John C. Jr., "Political Economy of Dodd-Frank: Why Financial Reform Tends to be Frustrated and Systemic Risk Perpetuated", *Cornell Law Review*, vol. 97, n.º 5, 2011, 1019-1082.

Coglianese, Cary/Lazer, David, "Management-based regulation: Prescribing private management to achieve public goals", *Law & Society Review*, 37, 4, 2003, 691-730.

Coglianese, Cary/Mendelson, Evan, "Meta-regulation and self-regulation", in *The Oxford Handbook of Regulation*, Oxford, Oxford University Press, 2010, 146–168.

Comissão Europeia, Livro Branco da Comissão Europeia sobre a inteligência artificial – Uma abordagem europeia virada para a excelência e a confiança, 2020 (available at https://op.europa.eu/pt/publication-detail/-/publication/ac957f13-53c6-11eaaece-01aa75ed71a101aa75ed71a1.0004.02/DOC_1&format=PDF)

Cunningham, Lawrence A., "The Appeal and Limits of Internal Controls to Fight Fraud, Terrorism, Other Ills", *The Journal of Corporation Law*, vol. 29, 2004, 267-336

Deloitte, "The case for artificial intelligence in combating money laundering and terrorist financing. A deep dive into the application of machine learning technology", 2018, 1-37.

Der Elst, Christoph and Van Daelen, Marijn, "Risk Management in European and American Corporate Law", in ECGI-Law Working Paper, No. 122, 2009

Domingos, Pedro, "The master algorithm: How the quest for the ultimate learning machine will remake our world", Basic Books, 2015.

EBA Report on automation in financial advice, 2016, (available at https://esasjointcommittee.europa.eu/Publications/Reports/EBA%20BS%202016%20422%20(JC%20SC%20CPFI%20Final%20Report%20on%20automated%20advice%20tools).pdf)

EBA Report on big data and advanced analytics, 2020 (available at https://www.eba.europa.eu/sites/default/documents/files/document_library//Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf)

EBA, EBA Analysis of Regtech in the EU Financial Sector, 2021

EBA, Guidelines on Internal Governance (EBA/GL/2017/11, of March 21st, 2018, available at https://www.eba.europa.eu/sites/default/documents/files/documents/10180/2164689/151a6ca-3-31ae-40b0-9f55-9d6c65b86b00/Guidelines%20on%20Internal%20Governance%20%28EBA-GL-2017-

EBA, Study of the Cost of Compliance with supervisory reporting requirements, 2021 (Report EBA/Rep/2021/15)

EBF, Position paper on AI in the banking industry, 2019, (available at https://www.ebf.eu/wp-content/uploads/2020/03/EBF_037419-Artificial-Intelligence-in-the-banking-sector-EBF.pdf)

Enriques, Luca/Zetzsche, Dirk A., "Corporate Technologies and the Tech Nirvana Fallacy", *European Corporate Governance Institute* (ECGI), n.º 457, 2019, 1-51

ESMA – Joint Committee Final Report on Big Data, 2018 (available at https://www.esma.europa.eu/sites/default/files/library/jc-2018-04_joint_committee_final_report_on_big_data.pdf)

Estrada, Juan Carlos, "The AML Arms Race: How Artificial Intelligence and Machine Learning Will Combat Money Laundering", in *Rutgers Bus. LJ*, 16, 2020, 383-408

European Banking Federation, "EBF position paper on AI in the banking industry", 2019, 1-42.

European Commission, "FinTech Action Plan: For a more competitive and innovative European financial sector", 2018

Fanto, James A., "The Professionalization of Compliance: Its Progress, Impediments, and Outcomes", *Notre Dame Journal of Law, Ethics & Public Policy*, vol. 35, n.º 1, 2021, 183-240

Fein, Melanie L., "How Should Robo-Advisors Be Regulated? Unanswered Regulatory Questions", in Allianz Global Investors, 2017, 1-14

Financial Conduct Authority, "Call for Input: Supporting the development and adoption of RegTech", https://www. fca. org. uk/publication/call-for-input/regtech-call-for-input. pdf, 2015

Fonseca, Patrícia Afonso, "As Novas Orientações da EBA em Matéria de Governo Interno", *A Emergência e o Futuro do Corporate Governance em Portugal*, vol. II, Almedina, Coimbra, 2018, 235-254

Fruth, Joshua, "Anti-money laundering controls failing to detect terrorists, cartels, and sanctioned states", March, 2018.

Gadinis, Stavros/Miazad, Amelia, "The Hidden Power of Compliance", *Minnesota Law Review*, vol. 103, 2019, 2135-2209.

Garrett, Brandon L, *Too Big to Jail*, Harvard University Press, Cambridge, 2014

Garrett, Brandon L. / Mitchell, Gregory, "Testing Compliance", in *Law and Contemporary Problems*, Vol. 83, No. 4, 2020, 47-84.

Gebauer/Niermann, in Hauschka/Moosmayer/Lösler Corporate Compliance, 3. Auflage, 2016.

Giuffrida, Iria, "Liability for AI Decision-Making: Some Legal and Ethical Considerations", *Fordham Law Review*, vol. 88, 2019, 439-456.

Griffith, Sean J., "Corporate Governance in an Era of Compliance", *William & Mary Law Review Online*, vol. 57, n.º 6, 2016, 2075-2140.

Gunnar Groh, *Creifelds kompakt, Rechtswörterbuch*, 4. Auflage, 2021, Beck-online

Hauschka/Moosmayer/Lösler, *Corporate Compliance*, 3. Auflage, 2016, Beck-online

Hess, David, "Ethical Infrastructure and Evidence-Based Corporate Compliance and Ethics Programs: Policy Implications from the Empirical Evidence", *New York University Journal of Law and Business*, Vol. 12, 2015.

Johnson, Kristin/Pasquale, Frank/Chapman, Jennifer, "Artificial intelligence, machine learning, and bias in finance: toward responsible innovation", *Fordham L. Rev.*, 88, 2019, 499-528.

Jung, John Ho Hee, "RegTech and SupTech: the future of compliance*", FinTech - Law and Regulation*, Elgar Financial Law and Practice, United Kingdom, 2019, 255-279

JWG, "Out of the window: COVID-19 prompts unexpected regulatory change for 2020 compliance, risk management work plans", 2020 (available at https://www.corlytics.com/newsreleases/out-of--the-window-covid-19-prompts-unexpected-regulatory-change--for-2020-compliance-riskmanagement-)

Kaya, Orçun, "Artificial intelligence in banking: A lever for profitability with limited implementation to date", in Deutsche Bank Research, 2019, 1-9.

Kingston, John, "Using artificial intelligence to support compliance with the general data protection regulation", *Artificial Intelligence and Law*, vol. 25, n.º 4, 2017, 429-443.

Kpmg, "There's a revolution coming. Embracing the challenge of RegTech 3.0", 2018, 1-12

Kurum, Esman, "RegTech solutions and AML compliance: what future for financial crime?", *Journal of Financial Crime*, ahead-of--print, ahead-of-print, 2020.

Labareda, João, "Contributo para o estudo do sistema de controlo e da função de cumprimento ("Compliance")", *Direito dos Valores Mobiliários*, 2016, 279-374.

Langevoort, Donald C, "Cultures of compliance", *American Criminal Law Review*, vol. 54, 2017, 933-977.

Langevoort, Donald C., "Monitoring: the behavioral economics of inducing agents' compliance with legal rules", *Georgetown Univer-*

sity Law Center Business, Economics and Regulatory Policy, Law and Economics Research Paper, n.º 276121, 2001, 1-39.

Lee, Joseph, "Access to Finance for Artificial Intelligence Regulation in the Financial Services Industry", *European Business Organization Law Review*, vol. 21, 2020, 731-757.

Lin, Tom C. W., "Artificial Intelligence, Finance, and the Law", *Fordham Law Review*, vol. 88, 2019, 531-551.

Lin, Tom C. W., "Compliance, Technology, and Modern Finance", *Brook. J. Corp. Fin. & Com. L.*, vol. 11, 2016, 159-181.

Lipshaw, Jeffrey M., "The False Dichotomy of Corporate Governance Platitudes", *The Journal of Corporation Law*, vol. 46, n.º 2, 2021, 346-384.

Lösler, Thomas, „Das moderne Verständnis von Compliance im Finanzmarktrecht", *NZG*, 2005, 104-108.

Machine learning in UK services, 2019, issued by the Bank of England and the FCA (available at https://www.bankofengland.co.uk/report/2019/machinelearning-in-uk-financial-services),

Magnuson, William, "A Unified Theory of Data", *Harvard Journal on Legislation*, vol. 58, 2021, 24-67.

Magnuson, William, "Artificial Financial Intelligence", *Harvard Business Law Review*, vol. 10, 2020, 338-382.

Maia, Pedro, "A robotização do mundo financeiro: reflexões introdutórias", in Estudos de Direito do Consumidor, n.º 16, Centro de Direito do Consumo - Instituto Jurídico, Coimbra, 2020, 273-306

Maia, Pedro, "Direito das Sociedades Bancárias", *Revista de Legislação e de Jurisprudência*, Ano 149º, n.º 4023, 2020, 372-411.

Marc Andreessen, "Why software is eating the world", *Wall Street Journal*, 2011.

Marco, Lamandini/Munoz David, Ramos, "A brief history of the evolution of financial institutions and of their regulation", *EU Financial Law. An introduction,* Cedam, Padova, 2016, 3-85.

Martinez, Veronica Root, "The Compliance Process", *Indiana Law Journal*, vol. 94, 2019, 203-251.

Maxwell, Winston J/Bourreau, Marc, "Technology neutrality in internet, telecoms and data protection regulation", *Computer and Telecommunications Law Review*, 31, 2014, 1-8.

Mayer, Colin, "The future of the corporation: Towards humane business", *Journal of the British Academy*, vol. 6, n.º 1, 2018, 1-16.

Mcneece, John B., "The Ethical Conflicts of the Hybrid General Counsel and Chief Compliance Officer", *Georgetown Journal of Legal Ethics*, vol. 25, 2012, 677-681.

Miller, Geoffrey P., "Risk Management and Compliance in Banks: The United States and Europe", *European Banking Union*, Oxford, United Kingdom, 2015, 200-216.

Miller, Geoffrey P., "The Role of Risk Management and Compliance in Banking Integration", *NYU Law and Economics Research Paper*, 2014, 14-34.

Miller, Geoffrey Parsons, "Compliance: Past, Present and Future", *University of Toledo Law Review*, vol. 48, 2016.

OECD, "OECD Council Recommendation on Artificial Intelligence", 2018 (available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449), adopted by the G20 in 2019 (available at https://www.mofa.go.jp/files/000486596.pdf).

Oliveira, Arlindo, *The Digital Mind: How Science is Redefining Humanity*, MIT Press, Cambridge, 2017.

Omarova, Saule T., "New Tech v. New Deal: Fintech as a Systemie Phenomenon", *Yale Journal on Regulation*, vol. 36, 2019, 735-793.

Orozco, David, "A Systems Theory of Compliance Law", *University of Pennsylvania Journal Business Law*, vol. 22, n.º 2, 2020, 244--302.

Packin, Nizan Geslevich, "RegTech, Compliance and Technology Judgment Rule", *Chicago-Kent Law Review*, vol. 93, n.º 1, 2018, 193-218.

Parker, Christine, "Meta-Regulation: Legal Accountability for Corporate Social Responsibility?", *The New Corporate Accountability: Corporate Social Responsibility and the Law*, Cambridge University Press, Cambridge, 2007, 207-237.

Parker, Christine, "The Open Corporation: Effective self-regulation and Democracy", Cambridge University Press, Cambridge, 2002

Piri, Michael M., "The Changing Landscapes of FindTech and RegTech: Why the United States Should Create a Federal Regulatory Sandbox", *Business & Finance Law Review*, vol. 2, n.º 2, 2019, 233-255.

Portugal Finlab Report, 2nd Edition, 2020 (available at https://8080dd92-d6fc-49d9-a97eb24c8f013bb2.filesusr.com/ugd/ca9a53_217c4187d5bd4a5a9b377c6f6500e0ff.pdf)

Rock, Edward B., "For Whom is the Corporation Managed in 2020?: The Debate over Corporate Purpose", in European Corporate Governance Institute - Law Working Paper, n.º 515, 2020, 20-16.

Rodrigues, Anabela Miranda, *Direito Penal Económico: Uma Política Criminal na Era Compliance*, 2.ª ed., Almedina, Coimbra, 2021.

Schneider, Uwe, "Compliance als Aufgabe der Unternehmensleitung", *ZIP*, 2003, 645-650.

Schumpeter, Joseph, *Capitalismo, Socialismo e Democracia*, Actual Editora, Coimbra, 2018.

Scopino, Gregory, "Key Concepts: Algorithms, Artificial Intelligence, and More", in *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and Other Derivatives*, Cambridge University Press, Cambridge, 2020, 13-47.

Scott, Colin, "Regulating everything: From mega-to meta-regulation", in Administration, vol. 60, 2012, 57-85

Simmons, Omari Scott/Dinnage, James D., "Innkeepers: A Unifying Theory of the In-House Counsel Role", in Seton Hall Law Review, vol. 41, n.º 1, 2011, 77-152.

Sokol, D. Daniel, "Twenty-Eighth Annual Corporate Law Center Symposium: Rethinking Compliance", *University of Cincinnati Law Review*, vol. 84, n.º 2, 2016, 399-420.

Solow-Niederman, Alicia, "Administering Artificial Intelligence", in S. Cal. L. Rev., 93, 2019, 633-696.

Sousa, Susana Aires De, "A colaboração processual dos entes coletivos: legalidade, oportunidade ou "troca de favores"?", in *Revista do Ministério Público*, n.º 158, 2019, 9-36.

The Global City, 2021, "2021: A Critical Year of RegTech", 2021, 1-76.

Treleaven, Philip, "Financial regulation of FinTech", *Journal of Financial Perspectives*, 3, 3, 2015, 1-14.

Van Der Elst, Christoph/Van Daelen, Marijn, "Risk Management in European and American Corporate Law", in ECGI-Law Working Paper, n.º 122, 2009.

Weber-Rey, Daniela, "Der Aufsichtsrat in der europäischen Perspektive - Vorschläge und Ideen für eine wirksame Corporate Governance", in NZG, 2013, 766-770

Yampolskiy, Roman V., "Unpredictability of AI", in Cornell University, 2019, 1-10.

Yang, Yueh-Ping (Alex) and Tsang, Chengyun, "RegTech and the New Era of Financial Regulators: Envisaging More Public-Private-Partnership Models of Financial Regulators", *University of Pennsylvania Journal of Business Law*, 2018, 354-404 (available at https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52021PC0206)

Zimiles, Ellen, "How AI is transforming the fight against money laundering", *World Economic Forum*, 2019 (available at https://www.weforum.org/agenda/2019/01/how-ai-can-knock-thestarch--out-of-money-laundering)

2

Algo-trading

(https://doi.org/10.47907/livro2021_4c3)

Alexandre de Soveral Martins[1]

Abstract:
A negociação algorítmica e, em especial, a de alta frequência, trouxe consigo riscos e vantagens. Os desenvolvimentos tecnológicos, permitindo que a informação circule a grande velocidade e seja também aproveitada muito rapidamente, podem contribuir para a instabilidade dos mercados financeiros. A Grande Crise Financeira de 2007/2008 (GFC) criou uma janela de oportunidade para que uma regulação mais apertada surgisse. Na União Europeia muitas alterações foram introduzidas a partir, sobretudo, de 2014. O enquadramento jurídico daí resultante foi, porém, visto de forma diferente por muitos atores, sendo de prever que algumas alterações surjam em cima da mesa em breve sob o efeito de aceleração do Brexit.

Keywords: negociação algorítmica; negociação de alta frequência; supervisão; regulação; mercados financeiros

I. INTRODUÇÃO. O PANORAMA

Segundo dados da ESMA fornecidos em 2014[2], na Euronext Lisbon 65% das ordens recebidas eram já de alta frequência. Nada de

[1] Associate Professor of Law. Univ. Coimbra, University of Coimbra Institute for Legal Research, Fac. Law.

[2] *Economic Report. High-frequency trading activity in EU equity markets*, Number 1, 2014, p. 11.

espantar no mundo financeiro, em que as novas tecnologias vão ganhando cada vez mais importância. De tal forma que se fala hoje da FinTech, expressão criada a partir das palavras *Financial Technology*[3].

A negociação algorítmica e, em particular, de alta frequência, ganha especial importância no quadro da negociação de futuros. No mercado de futuros dos EUA, mais de 90% da negociação era eletrónica em 2012[4]. Mas também no mercado a contado de valores mobiliários a negociação algorítmica e de alta frequência têm muita importância[5].

As novas realidades trazem consigo novos riscos. No início da tarde do dia 6 de maio de 2010 o *Dow Jones Industrial Average* caiu 998,5 pontos em 45 minutos, começando depois a recuperar e chegando ao fim do dia com apenas uma perda de 348 pontos. Em menos de cinco minutos, o *E-mini Standard and Poors 500 stock index futures contracts* caiu 5 pontos. Também se verificou grande volatilidade em relação a algumas ações. Mais tarde, percebeu-se que reações de algoritmos a uma ordem de venda muito grande de um fundo estariam subjacentes ao fenómeno[6].

Em 2014, no dia 6 de fevereiro, houve também um *flash-crash* do *Dax-Future*, que perdeu em segundos quase duzentos pontos[7]. Em 2018, no dia 5 de fevereiro, houve novamente um *flash-crash* com o índice do *Dow Jones Industrial Average*, que desceu de repente 1.597 pontos. No final do dia, ainda registava uma queda de 1175 pontos[8].

Porém, essas descidas e subidas muito rápidas, que ameaçam a estabilidade dos mercados pela extrema volatilidade que trazem consigo,

---

[3] Michael McGowan, «The rise of computerized high frequency trading: use and controversy», *Duke Law & Technology Review*, 9, 2009, s/p., lembra que a utilização de computadores na negociação em bolsa começou nos anos 70 do séc. XX com o sistema Designated Order Turnaround (DOT) na New York Stock Exchange.

[4] Federal Register, vol. 78, n.º 177, September 12, p. 56545 (disponível em https://www.federalregister.gov/documents/2013/09/12/2013-22185/concept-release-on-risk-controls-and-system-safeguards-for-automated-trading-environments).

[5] V. Michael McGowan, «The rise of computerized high frequency trading: use and controversy», *Duke Law & Technology Review*, 9, 2009, s/p..

[6] Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, CUP, Cambridge/New York/Port Melbourne/New Delhi, 2020, p. 413.

[7] Ralph Temporale, *Europïsche* Finanzmarktregulierung, Schäffer-Peowschel, Stuttgart, 2015, p. 57

[8] Dados recolhidos em Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 417.

não são os únicos perigos que podem ser identificados. A introdução de grandes números de ordens podem levar a que os sistemas de negociação colapsem[9]. A existência de mercados interconectados aumenta o risco de consequências sistémicas[10].

II. ALGUMAS REAÇÕES. TRAÇOS DO ENQUADRAMENTO JURÍDICO MAIS RELEVANTE

Da reunião do G 20 de Pittsburgh, em 2009, no seguimento da GFC de 2007/2008, tinha já emergido o objetivo de combater a instabilidade financeira[11], servindo de inspiração à Lei *Dodd-Franck*, nos EUA, que, embora tratando sobretudo de outros temas[12], também se ocupou do *spoofing*[13]. Como veremos melhor adiante, a HFT pode gerar instabilidade financeira.

A intervenção da União Europeia foi especialmente relevante[14]. Destacamos a DMIF II (Diretiva 2014/65/UE), de 15 de maio de 2014, relativa aos mercados de instrumentos financeiros[15], o RMIF

---

[9] Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 421.

[10] Referindo-se a um novo risco sistémico que resulta de participantes e produtos «too linked to fail», Tom Lin, «The New Market Manipulation», in 66 EmoryLJ, 2017, 1253-1314, a p. 1275.

[11] Johannes Karremans/Magnus Schoeller, «MiFID II between European rule-making and national market surveillance: the case of high-frequency trading», in Adrienne Héritier/Magnus Schoeller (ed.), *Governing Finance in Europe*, Elgar, Cheltenham/Northampton, 2020, p. 32 a 51, p. 35.

[12] V., p. ex., Randall Kroszner/Robert Shiller, *Reforming U.S. Financial Markets*. Reflections Before and Beyond Dodd-Frank, MIT Press, Cambridge, Massachusetts/London, 2011

[13] Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 349 e ss. e p. 450.

[14] Na França e na Alemanha já existia legislação importante desde, respetivamente, 2012 e 2013, embora a última tenha maior proximidade com as soluções da DMIF II: Johannes Karremans/Magnus Schoeller, «MiFID II between European rule-making and national market surveillance: the case of high-frequency trading», cit., a p. 37, salienta isso mesmo.

[15] A Proposta da Comissão, no entanto, já era de 2011: v. Proposta de Directiva do Parlamento Europeu e do Conselho relativa aos mercados de instrumentos financeiros, que revoga a Directiva 2004/39/CE do Parlamento Europeu e do Conselho (reformulação), COM(2011) 656 final. A DMIF I só tratava da negociação eletrónica de forma muito geral: v. Emanuel de Fournoux et al., *A new framework for European financial markets*, LexisNexis, London, 2019, p. 88.

(Regulamento 600/2014), o Regulamento Delegado 2017/565, de 25 de abril, o Regulamento Delegado 2017/578, de 13 de junho de 2016, sobre criação de mercado, o Regulamento Delegado 2017/589, de 19 de julho de 2016, sobre negociação algorítmica, o Regulamento Delegado 2017/588, de 14 de julho de 2016, sobre variações das ofertas de preços (tick-size), o MAR (*Market Abuse Regulation*, ou Regulamento 596/2014), o Regulamento Delegado 2016/522, de 17 de dezembro, v.g. sobre indicadores de manipulação de mercado e a Diretiva MAD (Market Abuse Directive, ou Diretiva 2014/57/EU, de 16 de abril de 2014, sobre as sanções penais aplicáveis ao abuso de informação privilegiada e à manipulação de mercado (abuso de mercado). As Q&A da ESMA também são importantes, bem como as suas *Guidelines* e os seus *Technical Standards*.

No ordenamento jurídico português, o CVM contém várias disposições que se ocupam da negociação algorítmica. São particularmente dignos de nota os arts. 317.º-E, 317.º-F, 317.º-G e 317.º-H, mas não podem igualmente deixar de ser referidos os arts. 208.º-A, 223.º-A, 311.º, 379.º, 1 e 2, 397.º-A e 400.º.

III. DEFINIÇÕES RELEVANTES

1. Negociação algorítmica

De acordo com o art. 4.º, 1, 39), da DMIF II, negociação algorítmica é «negociação em instrumentos financeiros, em que um algoritmo informático determina automaticamente os parâmetros individuais das ordens, tais como o eventual início da ordem, o calendário, o preço ou a quantidade da ordem ou o modo de gestão após a sua introdução, com pouca ou nenhuma intervenção humana. Esta definição não inclui qualquer sistema utilizado apenas para fins de encaminhamento de ordens para uma ou mais plataformas de negociação, para o processamento de ordens que não envolvam a determinação de parâmetros de negociação ou para a confirmação das ordens ou o processamento pós-negociação das transações executadas»[16]. Na definição sucinta de Matthias Lehmann, trata-se da utilização de *software*

[16] V. tb. o art. 317.º-E, 7, CVM.

para computador que determina automaticamente se, quando e onde a ordem é colocada[17].

Como facilmente se conclui a partir da definição da DMIF II, os encaminhadores de ordens automatizados (*Automated Order Routing* – AOR) parecem estar excluído da definição de negociação algorítmica, mas já poderá ser discutido se os encaminhadores de ordens inteligentes (*Smart Order Routing* – SOR), ao otimizarem o processo de execução de ordens, também estará[18]. Porém, o Considerando (22) do Regulamento Delegado 2017/565 considera que a «negociação algorítmica deve incluir encaminhadores de ordens inteligentes […] se esses dispositivos utilizarem algoritmos paa a otimização dos processos de execução de ordens que determinem os parâmetros da ordem para além da ou das plataformas através das quais a ordem será apresentada»[19]. Mais clara parece ser a exclusão do *post-trading* (*settlement*, *clearing*).

Tratando-se de uma empresa que prossegue uma estratégia de criação de mercado, aquela ficará sujeita à aplicação da DMIF II se preencher certas condições. O art. 2.º da DMIF II contém isenções. Porém, o seu n.º 1, d), exclui da isenção nele prevista os criadores de mercado. O art. 4.º, 1, 7), dá a seguinte definição de «Criador de mercado»: «uma pessoa que se apresenta nos mercados financeiros, com caráter contínuo, como estando disposta a negociar por conta própria através da compra e venda de instrumentos financeiros com base no seu próprio capital a preços que a própria define». De acordo com o Considerando (59), uma «empresa de investimento que desenvolve negociação algorítmica na prossecução de uma estratégia de criação de mercado deverá efetuar essa criação de mercado continuamente

---

[17] Matthias Lehmann, «Article 4 MiFIDII», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, Beck-Hart-Nomos, München/Oxford/Baden-Baden, 2019, p. 31.

[18] Telma Filipa Batista Gonçalves, «Estudo sobre os desafios da negociação algorítmica e de alta frequência na eficiência financiera e na integridade do mercado – novos desenvolvimentos regulatórios», *Direito dos Valores Mobiliários II*, ebook, IVM/AAFDL, 2018, p. 349, considera que no *Smart Order Routing* ainda haverá negociação algorítmica por haver seleção de plataformas de negociação e otimização de parâmetros de ordens (tempo, volume, preço).

[19] V. tb. o art. 18.º do Regulamento Delegado 2017/565. V., porém, o Considerando (27) e o art. 20.º, 2, do Regulamento Delegado referido.

durante uma proporção específica do horário de negociação da plataforma de negociação»[20].

A especial preocupação com os criadores de mercado compreende-se bem se pensarmos que um criador de mercado é «typically, today, an HFT posting buy and sell limit orders»[21]. A necessidade de negociar obriga o criador de mercado a procurar informação. Desde logo, a partir das ofertas de sentido contrário. O preço de mercado tenderá a refletir a informação e o criador de mercado irá atualizar os preços para compra e venda à medida que recolhe essas informações[22].

A negociação algorítmica pressupõe que a definição dos elementos das ordens introduzidas nos mercados tem pouca ou nenhuma intervenção humana. A Inteligência Artificial (IA) passa, por isso, a ter papel fundamental[23]. Mas o que é a IA? Haverá tantas definições como as de inteligência[24]. Inteligência Artificial pode ser apresentada como a utilização de computadores que realizam atividades que necessitariam de inteligência se fossem realizadas por seres humanos[25]. Um algoritmo, por sua vez, é um conjunto de instruções que determinam a atividade de um computador.

Há *software* que evolui ao treinar com exemplos, sem necessidade de nova intervenção dos programadores e levando a que se afirme que os computadores aprendem. Os algoritmos detetam constantes e atuam em conformidade, servindo até para fazer previsões. Quanto

---

[20] Lê-se ainda no mesmo Considerando que é «necessário clarificar, por meio de normas técnicas de regulamentação, o que se entende por uma proporção específica do horário de negociação da plataforma de negociação, assegurando que essa proporção específica seja significativa em comparação ao horário total de negociação, tendo em conta a liquidez, a dimensão e a natureza desse mercado específico e as características dos instrumentos financeiros negociados».

[21] Merritt Fox, «MiFID II and equity trading. A US View», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and MiFIR*, OUP, Oxford/New York, 2017, p. 487-525, a p. 489.

[22] Merritt Fox, «MiFID II and equity trading. A US View», cit., p. 493.

[23] V., p. ex., os OECD Principles on Artificial Intelligence, de 2019, o Relatório do *High Level Expert Group on Artificial Intelligence sobre Policy and investment recommendations for trustworthy AI*, de 2019, e, do mesmo grupo e também de 2019, as *Ethics Guidelines for trustworthy AI*.

[24] Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 19.

[25] Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 4.

maior a quantidade de dados (os *Big Data*) e quanto maior for a capacidade de processamento, mais rapidamente aprendem, podendo já produzir novos algoritmos.

Hoje, fala-se em Algo Bots, diminutivo de algoritmo e de robot, usados nos mercados financeiros para recolher dados, trabalhar esses dados e atuar em conformidade, comprando, vendendo ou mantendo. Os dados não são apenas os recolhidos a partir das operações realizadas no mercado. Dados sobre o PIB, desemprego, inflação, etc., podem ser tratados[26].

Como vimos, a definição de negociação algorítmica contida no art. 4.º, 1, 39), da DMIF II exige que se trate de negociação «com pouca ou nenhuma intervenção humana». O art. 18.º do Regulamento Delegado 2017/565 veio desenvolver a definição, estabelecendo que «deve considerar-se que um sistema tem pouca ou nenhuma intervenção humana sempre que, para qualquer processo de geração de ordens ou oferatas de preço ou para qualquer processo de geração de ordens ou ofertas de preço ou para qualquer processo de otimizaçaõ da execução de ordens, um sistema automatizado toma decisões em qualquer uma das fases de abertura, geração, encaminhamento ou execução de ordens ou ofertas de preços em conformidade com parâmetros predeterminados».

2. Negociação de alta frequência (HFT ou *High Frequency Trading*)

O art. 4.º, 1, 40), DMIF II considera HFT «uma técnica de negociação algorítmica caracterizada por: a) Uma infraestrutura destinada a minimizar a latência de rede e de outros tipo, incluindo pelo menos um dos seguintes sistemas para a entrada de ordens algorítmicas: partilha de instalações (co-location), alojamento de proximidade ou acesso eletrónico direto de alta velocidade; b) A determinação pelo sistema da abertura, geração, encaminhamento ou execução de ordens sem intervenção humana para as transações ou ordens individuais; e c) Elevadas taxas de mensagens intradiárias constituídas por ordens, ofertas de preços ou cancelamentos»[27].

---

[26] Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 46.

[27] Sobre a negociação algorítmica, v. os arts. 17.º, 48.º, 49.º e 50.º da DMIF II. V. tb. o art. 317.º-F CVM.

A HFT é, assim, uma modalidade da negociação algorítmica. Caracteriza-se, antes de mais, por uma infraestrutura destinada à minimização de latência, que consiste na «redução do intervalo de tempo entre a transmissão e a recepção de dados informáticos, o que implica o aumento da velocidade de propagação da informação, bem como a melhoria dos respectivos meios de transferência»[28]. O objetivo de minimizar a latência tem como razão de ser a necessidade de aumentar a capacidade de reação perante alterações que entretanto surjam[29].

Atualmente, a negociação já é realizada em milisegundos e, até, em microsegundos. Estar à frente dos outros pode significar a possibilidade de ganhar muito dinheiro. Para isso, é necessário aceder primeiro à informação relevante.

A infraestrutura destinada a minimizar a latência deve incluir «pelo menos um dos seguintes sistemas para a entrada de ordens algorítmicas: partilha de instalações (co-location), alojamento de proximidade ou acesso eletrónico direto de alta velocidade».

A co-locação é partilha de instalações (alojamento dos servidores nas instalações das plataformas de negociação)[30]. Pode ser especialmente perigosa se conferir acesso às bases de dados da plataforma de negociação antes de a informação ser pública[31]. Como se lê no Considerando (62) da DMIF II, «é fundamental exigir que as plataformas de negociação prestem esses serviços de partilha das instalações de forma não discriminatória, equitativa e transparente». Assim, o art. 48.º, 8, da DMIF II estabelece que os «Estados-Membros exigem que os mercados regulamentados garantam que as suas regras em matéria de

---

[28] José Manuel Quelhas, «High-frequency trading (HFT)», BCE, LVIII, 2015, p. 369 e ss., a p. 372. Por sua vez, Telma Filipa Batista Gonçalves, «Estudo sobre os desafios da negociação algorítmica e de alta frequência na eficiência financiera e na integridade do mercado – novos desenvolvimentos regulatórios», cit., p. 293, caracteriza a latência como «o tempo que uma mensagem leva para atravessar os canais de transmissão das ordens de mercado». Sobre a dependência da HFT de «ultra-low latency», Michael McGowan, «The rise of computerized high frequency trading: use and controversy», *Duke Law & Technology Review*, 9, 2009, s/p..

[29] V., p. ex., Miguel Santos Almeida, «High-frequency trading – Regulação e compliance no contexto da nova DMIF II», in Paulo Câmara (coord.), *O novo direito dos valores mobiliários*, Almedina, Coimbra, 2017, p. 427 e ss., a p. 431.

[30] Para exemplos, Michael McGowan, «The rise of computerized high frequency trading: use and controversy», *Duke Law & Technology Review*, 9, 2009, s/p..

[31] Telma Filipa Batista Gonçalves, «Estudo sobre os desafios da negociação algorítmica e de alta frequência na eficiência financiera e na integridade do mercado – novos desenvolvimentos regulatórios», cit., p. 329.

serviços de partilha das instalações sejam transparentes, equitativas e não discriminatórias».

A minimização da latência também pode ser conseguida através do alojamento de proximidade (*proximity hosting*), que consiste no uso de servidores de terceiros próximos das instalações da plataforma de negociação, e do acesso eletrónico direto de alta velocidade.

O acesso eletrónico direto consiste no «mecanismo através do qual um membro, participante ou cliente numa plataforma de negociação permite que uma pessoa utilize o seu código de negociação para que possa transmitir por via eletrónica diretamente à plataforma de negociação ordens relativas a um instrumento financeiro e inclui mecanismos que envolvam a utilização, por uma pessoa, da infra-estrutura do membro, participante ou cliente ou de qualquer sistema de conexão por ele disponibilizado para transmitir ordens (acesso direto de mercado) e os mecanismos em que essa infra-estrutura não seja utilizada por uma pessoa (acesso patrocinado)» – v. art. 4.º, 1, 41), da DMIF II.

O acesso eletrónico direto a plataformas de negociação está sujeito a uma série de limitações. O art. 17.º, 5, da DMIF II[32] ocupa-se disso mesmo, exigindo, designadamente, que a empresa de investimento que proporciona esse acesso disponha de «sistemas e controlos eficazes que assegurem a realização de uma avaliação e análise corretas da aptidão dos clientes que utilizam o serviço, que os clientes que utilizam o serviço estão impedidos de ultrapassar limiares de crédito e de negociação pré-estabelecidos e adequados, que a negociação por clientes que utilizam o serviço é devidamente acompanhada e que os controlos de risco adequados impedem que a negociação seja suscetível de criar riscos para a própria empresa de investimento ou de criar ou contribuir para perturbações no mercado ou ser contrário ao disposto no Regulamento (UE) n.º 596/2014 ou às regras da plataforma de negociação», sendo proibido aquele acesso sem estes controlos. São ainda de destacar os deveres de controlo, comunicação e registo ali estabelecidos.

É também importante o que se lê no art. 21.º, 4, do Regulamento Delegado 2017/589, de 19 de julho de 2016[33]: «Um prestador de DEA que permita a um cliente de DEA conceder o seu acesso de DEA

---

[32] V. tb. arts. 317.º-H e 397.º-A CVM.

[33] Trata-se do Regulamento que complementa a DMIF II «no que diz respeito às normas técnicas de regulamentação que especificam os requisitos em matéria de organização das empresas de investimento que realizam negociação algorítmica».

aos seus próprios clientes («subdelegação») deve estar apto a identificar os diferentes fluxos de ordens dos beneficiários dessa subdelegação sem ser obrigado a conhecer a identidade dos beneficiários desse acordo».

Como resulta da respetiva definição, na HFT a velocidade soma-se à intensidade da negociação, com elevadas taxas de mensagens intradiárias. Estas últimas «consistem no envio, em média, de um dos seguintes: a) Pelo menos duas mensagens por segundo no que diz respeito a um único instrumento financeiro negociado numa plataforma de negociação; b) Pelo menos quatro mensagens por segundo no que diz respeito a todos os instrumentos financeiros negociados numa plataforma de negociação»: v. o art. 19.º do Regulamento 2017/565.

A utilização de HFT leva a que não se possam usar várias isenções de autorização como empresa de investimento previstas no art. 2.º da DMIF II. Se negociarem por conta própria com HFT, são empresas de investimento e necessitam de autorização para atuarem como tal (art. 2.º, 1, d), iii), DMIF II; para a negociação em derivados de mercadorias ou licenças de emissão ou seus derivados, v. o art. 2.º, 1, j)). Os negociadores por conta própria que realizam negociação de alta frequência terão de ser autorizados como empresa de investimento, com as consequências daí resultantes no que diz respeito aos requisitos da Diretiva 2013/36/UE[34].

IV. ALEGADAS VANTAGENS E ALGUNS PERIGOS

Há quem diga que a negociação algorítmica torna os mercados mais racionais, pois afasta as emoções (medo, pânico, ira, ódio). É alegado que também aumenta liquidez, aumenta a eficiência na formação de preços e pode baixar spreads[35]. Os algoritmos aceleram os processos e economizam custos ao eliminarem a intervenção humana[36].

---

[34] Pierre-Henri Cognac, «Algorithmic Trading and High-Frequency Trading (HFT)», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and MiFIR*, cit., p. 469-485, a p. 482.

[35] Telma Filipa Batista Gonçalves, «Estudo sobre os desafios da negociação algorítmica e de alta frequência na eficiência financiera e na integridade do mercado – novos desenvolvimentos regulatórios», cit., p. 296.

[36] Gerald Spindler, «Controlo of Algorithms in Financial Markets. The Example of High-Frequency Trading», in Martin Ebers/Susana Navas (ed.), *Algorithms and Law*, CUP, Cambridge/New York/Port Melbourne/New Delhi, 2020, p. 207-220, a p. 207.

O Considerando (62) da DMIF II revela uma ponderação dessas vantagens. Destacamos a seguinte passagem: «As tecnologias de negociação proporcionaram de um modo geral benefícios ao mercado e aos participantes no mercado, tais como uma maior participação nos mercados, um aumento da respetiva liquidez, menores diferenciais, uma menor volatilidade a curto prazo e os meios para obter uma melhor execução das ordens dos clientes».

No entanto[37], a negociação algorítmica e, no âmbito desta, a de alta frequência envolvem riscos para a estabilidade do mercado. Muita da liquidez aparentemente gerada é falsa se as ordens são canceladas e as operações não se concluem. É também o resultado de perspetivas de curtíssimo prazo, acabando por afetar o processo de formação de preços[38].

O art. 311.º, 2, e), do CVM considera que são suscetíveis de pôr em risco a regularidade de funcionamento, a transparência e a credibilidade do mercado os «padrões de intervenção negocial algorítmica ou de alta frequência que comportem riscos de perturbação, de alteração artificial ou enganosa da negociação ou de atraso no funcionamento do sistema de negociação». A HFT cria, inclusivamente, o perigo de manipulação de mercado[39]. São destacados[40] os comportamentos que consistam em «ping orders», «quote stuffing», «momentum ignition», «layering and spoofing». Refere-se ainda[41] a utilização da HFT para se fazer «front running» e «slow market arbitrage», o que se torna mais fácil através da co-locação.

No Considerando (62) da DMIF II também se faz referência aos riscos a que as tecnologias de negociação dão origem: sobrecarga dos sistemas das plataformas de negociação, reação excessiva e agravamento

---

[37] V., p. ex., Pierre-Henri Cognac, «Algorithmic Trading and High-Frequency Trading (HFT)», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and MiFIR*, cit., p. 469-485, a p. 469,

[38] Telma Filipa Batista Gonçalves, «Estudo sobre os desafios da negociação algorítmica e de alta frequência na eficiência financiera e na integridade do mercado – novos desenvolvimentos regulatórios», cit., p. 297.

[39] V., p. ex., a decisão, citada por Pierre-Henri Cognac, da *Commission des Sanctions* da AMF no processo contra *Euronext Paris SA* e *Virtu Financial Europe Ltd*, de 4 de dezembro de 2015.

[40] Pierre-Henri Cognac, «Algorithmic Trading and High-Frequency Trading (HFT)», cit., p. 483.

[41] Merrit Fox, «18. MiFID II and Equity Trading: a US View», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and MiFIR*, cit., p. 487 e ss., a p. 499.

da volatilidade, manipulação de mercado e fuga para *dark markets* para não ter de interagir com negociantes que recorrem à HFT. Com efeito, o risco de rápidos movimentos (*flash-events*), ao aumentar a instabilidade e a volatilidade, dificulta a correta identificação do valor, o que pode aumentar a atratividade dos *dark markets*[42]. A possível atuação de *hackers* já foi também referida[43].

No art. 12.º, 2, c), do Regulamento 596/2014 (MAR ou *Market Abuse Regulation*) lê-se o seguinte: «Considera-se como manipulação de mercado, entre outros, a seguinte conduta: […] c) Colocar ordens numa plataforma de negociação, incluindo o seu cancelamento ou alteração, por meio de qualquer mecanismo de negociação, incluindo meios eletrónicos como estratégias de negociação algorítmica e de alta frequência, tendo um dos efeitos referidos no n.º 1, alíneas a) ou b), ao: i) perturbar ou atrasar o funcionamento do sistema de negociação da plataforma de negociação, ou que seja idónea para o fazer, ii) dificultar a identificação por outras pessoas de ordens verdadeiras no sistema de negociação da plataforma de negociação, ou que seja idónea para o fazer, nomeadamente através da introdução de ordens que resultem na sobrecarga ou desestabilização do livro de ofertas, ou iii) gerar, ou ser idónea para gerar, uma indicação falsa ou enganosa sobre a oferta ou a procura, ou o preço, de um instrumento financeiro, nomeadamente através da introdução de ordens para iniciar ou exacerbar uma tendência».

Por sua vez, as als. a) e b) do n.º 1 do art. 12.º do MAR dispõem que, para «efeitos do presente regulamento, manipulação de mercado engloba as seguintes atividades: a) Realizar uma operação, colocar uma ordem ou qualquer outra conduta que: i) dê, ou seja idónea para dar, indicações falsas ou enganosas no que respeita à oferta, à procura ou ao preço de um instrumento financeiro, de um contrato de mercadorias à vista com ele relacionado ou de um produto leiloado baseado em licenças de emissão, ou ii) assegure, ou seja idónea para assegurar, o preço de um ou mais instrumentos financeiros, de contratos de mercadorias à vista com eles relacionados ou de um produto leiloado baseado em licenças de emissão, a um nível anormal ou artificial; exceto se a pessoa

---

[42] Johannes Karremans/Magnus Schoeller, «MiFID II between European rule-making and national market surveillance: the case of high-frequency trading», cit., p. 35, Martin Konstantin Thelen, *Dark Pools*, Duncker & Humblot, Berlin, 2019, p. 27.

[43] Gerald Spindler, «Controlo of Algorithms in Financial Markets. The Example of High-Frequency Trading», cit., a p. 209,

que realizou as operações, colocou as ordens ou praticou outra conduta faça prova de que essa operação, ordem ou conduta tiveram lugar por razões legítimas e se encontram em conformidade com as práticas de mercado aceites, definidas nos termos do artigo 13.º; b) Realizar operações, colocar uma ordem ou qualquer outra atividade ou conduta que afete, ou seja idónea para afetar, o preço de um ou mais instrumentos financeiros, um contrato de mercadorias à vista com eles relacionado ou um produto leiloado baseado em licenças de emissão, recorrendo a procedimentos fictícios ou quaisquer outras formas de engano ou artifício […]». Se for concedido o acesso eletrónico direto para transmitir ordens à plataforma de negociação, é fácil ver o risco envolvido caso não existam ou falhem os controlos quanto a esse acesso.

V. ALGUNS COMPORTAMENTOS QUE PODEM SER FACILITADOS PELA HFT

A HFT pode ser utilizada no âmbito de comportamentos que podem constituir manipulação de mercado. Muitos desses comportamentos vêm referidos como indicadores de manipulação de mercado no Regulamento Delegado (UE) 2016/522, de 17 de dezembro de 2015, embora a lista nele contida não seja exaustiva nem determinante, como resulta do Considerando (6). Essa lista consta do respetivo Anexo II (Indicadores de manipulação).

Logo na Sec. 1, 1, c), surgem referidas as *ping orders*, que consistem em «pequenas ordens de negociação a fim de determinar o nível de ordens ocultas e, em especial, avaliar o que se encontra numa plataforma opaca». Essas pequenas ordens até podem ser executadas.

Por sua vez, a Sec. 1, 1, d), faz referência ao *phishing*, que é descrita como «execução de ordens de negociação, ou de uma série de ordens de negociação, a fim de descobrir ordens de outros participantes e, em seguida, emitir uma ordem de negociação com o intuito de tirar partidodas informações obtidas […]». No *phishing* são lançadas ordens ou séries de ordens para ver outras a aparecerem, aproveitando-se as informações que daí resultam para lançar outra ordem[44].

---

[44] V. Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 459, e Carsten Gerner-Beuerle, «Article 12 MAR», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, cit., p. 738.

Na Sec. 1, 4, e), vemos surgir referida a *quote stuffing*: «colocação de um grande número de ordens de negociação e/ou cancelamentos e/ou atualizações de ordens de negociação a fim de criar incerteza junto dos outros participantes, abrandando o seu processo, e/ou camuglar a sua própria estratégia […]». Trata-se, assim, da inserção de ordens muito variadas, cancelamentos e atualizações para criar instabilidade e ocultar o que se pretende e tornar difícil conhecer o livro de ofertas[45].

A *momentum ignition* vem referida na Sec. 1, 4, f): «colocação de ordens de negociação ou de uma série de ordens de negociação ou execução de operações ou de uma série de operações sucetíveis de iniciar ou exacerbar uma tendência e incentivar outros particcipantes a acelerar ou alargar a tendência a fim de criar uma oportunidade para encerrar ou abrir uma posição a um preço favorável […]. Esta prática pode igualmente ser ilustrada pelo elevado rácio de ordens canceladas (p. ex., rácio de ordens de negociação), que pode ser combinado com um rácio de volume (p. ex., número de instrumentos financieros por ordem)». É, assim, uma prática que pode ser usada para iniciar ou reforçar uma tendência e aproveitar-se dela, com introdução de elevado número de ofertas, desfazendo depois a posição. Se as ordens são canceladas antes da sua execução, teremos *spoofing*.

O *layering and spoofing* vem caracterizado na Sec. 1, 5, e): «apresentação de ordens de negociação múltiplas ou de grande dimensão, frequentemente inatingíveis, num lado do registo de ordens, a fim de executar uma negociação no outro lado do registo de ordens. Assim que a negociação é efetuada, as ordens sem intenção de ser executadas são retiradas […]». Também aqui se procura obter ganhos de um lado com o que se provocou do outro.

O *spoofing* consiste na apresentação de uma ordem com a intenção de a cancelar logo de seguida e aproveitar as reações para ganhar de outra forma. Com o *spoofing* são dadas informações sobre a procura ou oferta que não correspondem à verdade, pois há a intenção de cancelar essa mesma ordem. Procura-se saber como serão as reações a cada ordem e daí retirar conclusões.

---

[45] Telma Filipa Batista Gonçalves, «Estudo sobre os desafios da negociação algorítmica e de alta frequência na eficiência financiera e na integridade do mercado – novos desenvolvimentos regulatórios», cit., 310.

O *spoofing* pode ser utilizado em combinação com o *layering*, procurando ganhar com o preço que se induziu[46]. A atuação pode criar a impressão de que, por exemplo o aumento da procura indiciada levará a uma subida dos preços, conduzindo o mercado a agir em conformidade. Também pode ser criada impressão oposta, gerando a ideia de que há grandes quantidades para vender e provocando a descida do preço[47]. E o algoritmo pode ir alternando os movimentos. Como a reação às ordens é na casa dos milisegundos, as ordens podem ser canceladas muito rapidamente também.

Merece também referência o *wash trading*, que vem mencionado na Sec. 1, 3, a): «celebração de acordos de compra ou venda de um instrumento financeiro[…] em que não existe alteração de direitos de usufruto ou risco de mercado ou em que o direito de usufruto ou o risco de mercado é transferido entre partes que agem em concertação ou conluio […]. Esta prática pode igualmente ser ilustrada pelos seguintes indicadores de manipulação de mercado adicionais: i) repetição pouco habitual de uma operação entre um pequeno número de partes durante um determinado período, ii) operações ou ordens de negociação que alteram, ou são suscetíveis de alterar, a avaliação de uma posição sem diminuir/aumentar a dimensão da posição, iii) o indicador estabelecido no ponto 1, alínea a), subalínea i), da presente secção». Teremos muitas vezes compras e vendas combinadas sem risco económico ou alteração na posição líquida[48], sendo apresentadas ordens que, na verdade, têm preço combinado ou sem risco, existindo negociação consigo ou com entidade controlada, ou preços combinados (*matched orders*). Acabam por ser materialmente fictícias, dando a falsa impressão de atividade no mercado[49].

---

[46] Para descrições, Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 345, e Carsten Gerner-Beuerle, «Article 12 MAR», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, cit., p. 745.

[47] Para um exemplo, Telma Filipa Batista Gonçalves, «Estudo sobre os desafios da negociação algorítmica e de alta frequência na eficiência financiera e na integridade do mercado – novos desenvolvimentos regulatórios», cit., p. 309.

[48] V. Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, p. 329, e Carsten Gerner-Beuerle, «Article 12 MAR», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, cit., p. 739.

[49] Carsten Gerner-Beuerle, «Article 12 MAR», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, cit., p. 740)

Por seu lado, o *marking the close*, como se lê na Sec. 1, 5, d), consiste na «compra ou venda de um instrumento financeiro […] deliberadamente, no momento de referência da sessão de negociação (p. ex., abertura, encerramento, liquidação) num esforço para aumentar, diminuir ou manter o preço de referência (p. ex., preço de abertura, preço de encerramento, preço de liquidação) num nível específico […]. Esta prática pode igualmente ser ilustrada pelos seguintes indicadores de manipulação de mercado adicionais: i) colocação de ordens que representam volumes significativos no registo de ordens central do sistema de negociação alguns minutos antes da fase de determinação do preço do lielão e cancelamento de tais ordens alguns segundos antes do congelamento, do registo de ordens para efetios do cálculo do preço de licitação, para que o preço teórico de abertura possa parecer mais elevado/baixo do que de outra forma aconteceria, ii) os indicadores estabelecidos no ponto 1, alínea b), subalíineas i), iii), iv) e v), da presente secção, iii) realização de operações ou apresentação de ordens de negociação, nomeadamente perto de um ponto de referência durante o dia de negociação, que, devido à sua dimensão em relação ao mercado, terão claramente um impacto significativo na oferta ou procura, ou no preço ou valor, iv) operações ou ordens de negociação sem nenhuma outra justificação aparente que não a de aumentar/diminuir o preço ou aumentar o volume de negociação, nomeadamente perto de um ponto de referência durante o dia de negociação – p. ex., na abertura ou perto do encerramento». Ao comprar ou vender no final do dia de negociação pode estar subjacente uma estratégia destinada a influenciar os preços dos derivados.

Queremos ainda deixar uma referência ao s*moking*, mencionado na Sec. 1, 6, j): «colocação de ordens de negociação para atrair outros participantes no mercado através do recurso a técnicas tradicionais de negociação (*slow traders*) que são, em seguida, rapidamente revistas em condições menos vantajosas, na esperança de uma execução rentável em relação ao fluxo de entrada de ordens de negociação de *slow traders* […]»[50]. Quem usa a HFT pode querer gerar a impressão de que existe muita atividade num mercado para atrair os *slow traders*, colocando ordens a preços sucessivamente mais altos e depois

[50] Cfr. o Anexo II do Regulamento Delegado 2016/522, Secção 1, 6, j):

revendo as ordens em termos menos vantajosos e conseguindo a sua execução[51].

Poderá ainda verificar-se a obtenção de informação privilegiada se os algoritmos utilizados conseguem ler rapidamente a intenção de um investidor que pretende adquirir posição longa, antecipando-se nas plataformas e adquirindo a um preço mais baixo, para depois oferecer a um preço mais alto quando chega a ordem do investidor. Há o claro perigo de ocorrer *front running*, sendo tomada a decisão de investimento com base no conhecimento de uma ordem que pode afetar o preço de mercado, mas que não é do conhecimento do público[52] e ainda não consta do livro de ordens[53], aumentando o risco para a estabilidade dos mercados.

Os algoritmos já se vigiam uns aos outros[54], tentando vencer-se. Existe o perigo de vermos Algo Bots a iniciarem processos que influenciam comportamentos das pessoas e dos outros Algo Bots para terem recompensas. O Facebook já fez uma experiência em que colocou Bots a negociarem uns com os outros, com vista a alcançarem determinados objetivos. Verificou-se que os programas eram capazes de tentar enganar-se reciprocamente, simulando interesse em algo que não teria valor, para depois usarem isso em compromissos futuros[55]. Os algoritmos não correm o risco de ser presos[56]. E também não será fácil saber onde está o dolo. Isso remete-nos para um outro problema: o de saber que enquadramento jurídico dar às empresas de informática e aos matemáticos que concebem o *software*[57].

---

[51] Carsten Gerner-Beuerle, «Article 12 MAR», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, cit., p. 746 e s..

[52] Telma Filipa Batista Gonçalves, «Estudo sobre os desafios da negociação algorítmica e de alta frequência na eficiência financiera e na integridade do mercado – novos desenvolvimentos regulatórios», cit., p. 324.

[53] V. Carsten Gerner-Beuerle, «Article 12 MAR», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, cit., p. 737.

[54] V. Gerald Spindler, «Controlo of Algorithms in Financial Markets. The Example of High-Frequency Trading», cit., a p. 208.

[55] Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 295.

[56] V. a citação de Lynn Lopucki em Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. 296.

[57] Gregory Scopino, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, cit., p. xi.

VI. ALGUMAS MEDIDAS

1. Aspetos gerais

Sobre a negociação de alta frequência e a utilização de algoritmos, lê-se o seguinte no Considerando 64 da DMIF II: «As empresas de investimento e as plataformas de negociação deverão assegurar a aplicação de medidas rigorosas para garantir que a negociação algorítmica ou as técnicas de negociação algorítmica de alta frequência não criem perturbações no mercado e não possam ser utilizada para fins abusivos. As plataformas de negociação deverão igualmente assegurar que os seus sistemas de negociação são sólidos e foram adequadamente testados para lidar com maiores fluxos de ordens ou situações de tensão no mercado e que existem interruptores (circuit breakers) nas plataformas de negociação para parar ou restringir temporariamente a negociação se se verificarem subitamente movimentos de preços inesperados». Vemos aqui apresentado um conjunto vasto de propostas para abordar alguns dos problemas que a negociação algorítmica pode gerar. São, porém, ainda e apenas ideias gerais, cuja concretização importa conhecer.

Na DMIF II e no RMIF vamos encontrar a tentativa de limitar os riscos envolvidos na HFT através de vários regimes. Destacamos os que exigem autorizações, obrigam a cumprir deveres de organização e reporte, a manter documentação, a disponibilizar *circuit-breakers* e a respeitar o *tick-size* (variação mínima que pode ser executada).

2. O art. 17.º da DMIF II e o Regulamento Delegado (UE) 2017/589

Lê-se no art. 17.º, 1, da DMIF II que uma empresa de investimento que desenvolva negociação algorítmica fica sujeita a uma longa lista de deveres quanto aos seus sistemas e controlos de risco e aos sistemas de negociação que utiliza. Vemos igualmente ali estabelecidos deveres quanto a planos de continuidade das atividades e à testagem e acompanhamento dos sistemas[58].

Os demais números do art. 17.º da DMIF II são igualmente importantes. Ali encontramos prevista a necessidade de consagrar deveres de

[58] Muitos dos deveres que analisaremos podem dar origem a responsabilidade contraordenacional se não forem cumpridos: cfr. o art. 397.º do CVM.

comunicação e registo, um enquadramento para as empresas de investimento que desenvolvam negociação algorítmica para prosseguir uma estratégia de criação de mercado e, bem assim, para as que proporcionam acesso eletrónico direto ou atuem como membro de compensação geral para outras pessoas.

O Regulamento Delegado (UE) 2017/589, de 19 de julho de 2016, veio especificar os requisitos referidos no art. 17.º, 1, da DMIF II. Seguindo a ordem por que as matérias são tratadas naquele Regulamento, vemos surgirem, designadamente, requisitos gerais em matéria de organização, exigências quanto aos sistemas de negociação (testagem de sistemas, algoritmos ou estratégias, implantação), gestão pós-implantação, gestão de alterações, meios para assegurar a resistência (*kill functionality*, deteção de manipulações de mercado, planos de continuidade, controlos, segurança e limites no acesso), acesso eletrónico direto e registo das ordens na HFT. É ainda possível encontrar no mencionado instrumento preceitos relativos ao acesso eletrónico direto, à atuação das empresas de investimento como membros de compensação gerais e à HFT. Vejamos alguns desses regimes com mais atenção, procurando sublinhar os aspetos mais relevantes tendo em conta os problemas anteriormente analisados.

3. Testagem

O estabelecimento de obrigações de testagem é particularmente visível em várias disposições do Regulamento Delegado 2017/589[59]. Desde logo no seu art. 5.º, que consagra uma metodologia geral. Decorre do seu n.º 1 que as empresas de investimento devem, antes de implantar ou atualizar sistema de negociação algorítmica, algoritmo de negociação ou estratégia de negociação algorítmica, estabelecer metodologias para desenvolver e testar os sistemas, algoritmos ou estratégias. Essas metodologias devem dizer respeito à conceção, desempenho, conservação de registos e aprovação do sistema de negociação algorítmica, do algoritmo de negociação ou da estratégia de negociação algorítmica (n.º 3) e devem assegurar que o sistema de negociação algorítmica, o algoritmo de negociação ou a estratégia de negociação algorítmica não se comporta de forma não pretendida, cumpre as obrigações da empresa de investimento decorrentes do Regulamento, cumpre as regras

[59] V. tb. o art. 48.º, 6, da DMIF II.

e sistemas das plataformas de negociação acedidas, não contribui para condições irregulares de negociação, continua a funcionar eficazmente em condições de tensão do mercado e, se necessário, nessas condições, permite a desativação do sistema de negociação algorítmica ou do algoritmo de negociação (n.º 4).

Estão ainda previstos testes suplementares em caso de alterações substanciais ao sistema de negociação algorítmica ou acesso à plataforma de negociação em que sistema, algoritmo ou estratégia serão utilizados (art. 5.º, 5).

Não são apenas as empresas de investimento que a DMIF II quer ver a realizar testes aos algoritmos. Com efeito, o art. 48.º, 6, da DMIF II estabelece que os Estados-Membros exigem aos mercados regulamentados testes apropriados aos algoritmos para que os sistemas de negociação que os utilizam «não criam nem contribuem para a perturbação da negociação no mercado e para gerir quaiser perturbamções que afetem a negociação decorrentes desses sistemas de negociação algorítmica, incluindo sistemas que limitem o rácio de ordens não executadas face às transações que podem ser introduzidas no sistema por um membro ou participante, a fim de poder abrandar o fluxo de ordens, se se verificar o risco de ser atingida a capacidade máxima do sistema, e de limitar e fazer cumprir a variação mínima da oferta de preço (tick) que pode ser executada no mercado».

Voltando ao Regulamento Delegado 2017/589, verificamos ainda que no art. 6.º estão previstos testes de conformidade aos sistemas de negociação algorítmica e algoritmos de negociação.

Por sua vez, no art. 10.º surgem referidos testes de esforço a realizar pela empresa de investimento no âmbito da sua autoavaliação anual para verificar se os sistemas de negociação algorítmica estão aptos a suportar aumentos de fluxos de ordens ou tensões no mercado. Os testes devem incidir, nomeadamente, sobre a capacidade de processamento de elevados volumes de mensagens e de elevados volumes de negociação. Compreendem-se estes testes de esforço, pois as empresas de investimento devem realizar anualmente um processo de autoavaliação e validação. Nesse processo devem rever, avaliar e validar, v.g., os sistemas de negociação algorítmica, os algoritmos de negociação e as estratégias de negociação algorítmica (art. 9.º do Reglamento Delegado 2017/589).

4. Circuit breakers

Os *circuit breakers* (ou *kill functionality*) são importantes para evitar movimentos muito profundos num sentido ou noutro. Se os preços se alterarem mais do que x em relação a um determinado valor de referência, para cima ou para baixo, a negociação é suspensa ou encerra mais cedo.

O art. 17.º, 1, da DMIF II dispõe que a empresa de investimento que desenvolva negociação algorítmica deve adotar sistemas de negociação que não funcionem de mode a criar ou contribuir para uma perturbação de mercado.

No Considerando (9) do Regulamento Delegado 2017/589, de 19 de julho de 2016[60], pode ler-se também que as «empresas de investimento devem poder retirar a totalidade ou parte das suas ordens sempre que tal seja necessário (funcionalidade de interrupção ou «kill functionality»). Para que essa retirada seja eficaz, a empresa de investimento deve estar sempre em condições de identificar os algoritmos de negociação, os operadores ou os clientes que são responsáveis por uma dada ordem». É especialmente relevante o art. 12.º, 1, daquele Regulamento: «As empresas de investimento devem estar aptas a cancelar de imediato, como medida de emergência, uma ou a totalidade das suas ordens não executadas apresentadas a uma ou a todas as plataformas de negociação às quais a empresa de investimento se encontra conectada»[61].

---

[60] Trata-se do Regulamento que complementa a DMIF II «no que diz respeito às normas técnicas de regulamentação que especificam os requisitos em matéria de organização das empresas de investimento que realizam negociação algorítmica».

[61] No âmbito dos controlos pós-negociação, o Regulamento Delegado 2017/589 dispõe, no seu art. 17.º, que as «empresas de investimento devem operar de forma contínua os controlos pós-negociação que têm em vigor. Sempre que um controlo pós-negociação é iniciado, a empresa de investimento deve tomar as medidas adequadas, que podem incluir o ajustamento ou a desativação do algoritmo de negociação ou o sistema de negociação relevante, ou uma retirada ordenada do mercado». Por sua vez, e quanto aos mercados regulamentados, a DMIF II estabelece, no seu art. 48.º, 5, que os Estados-Membros devem exigir que os mesmos «possam interromper ou restringir temporariamente a negociação» em caso de variação significativa dos preços de um instrumento financeiro. V. tb. o art. 213.º-A CVM.

5. Variação de ofertas de preços (*tick sizes*). Ordens não executadas

No que diz respeito aos mercados regulamentados (e não só), o art. 48.º, 6, DMIF II dispõe que devem dispor de «sistemas, procedimentos e mecanismos eficazes […] para gerir quaisquer perturbações que afetem a negociação decorrentes desses sistemas de negociação algorítmica, incluindo sistemas que limitem o rácio de ordens não executadas face às transações que podem ser introduzidas no sistema por um membro ou participante, a fim de poder abrandar o fluxo de ordens, se se verificar o risco de ser atingida a capacidade máxima do sistema, e de limitar e fazer cumprir a variação mínima da oferta de preço (tick) que pode ser executada no mercado»[62].

Os *tick sizes* são objeto de tratamento desenvolvido no art. 49.º da DMIF II e no Regulamento Delegado 2017/588 de 14 de julho de 2016. Com aquele regime pretende-se obter um equilíbrio entre preços razoavelmente estáveis e a necessidade de evitar constrangimentos ao estreitamento dos *spreads*[63] relativamente aos instrumentos financeiros abrangidos. Um *tick size* demasiado pequeno pode aumentar a frequência de modificações e cancelamentos[64], o que poderia estimular os algoritmos. Por outro lado, um *tick size* muito grande pode afastar o interesse em apresentar ofertas[65].

Por sua vez, o Regulamento Delegado 2017/566 de 18 de maio de 2016 ocupa-se especialmente do rácio entre as ordens não executadas e as transações (*order to trade ratio* - OTR) de modo a evitar perturbações das condições de negociação. Trata-se de um tema particularmente importante tendo em conta que as ordens podem não ter sido executadas precisamente por se inserirem numa estratégia de manipulação de mercado. Pretende-se também evitar uma volatilidade excessiva, como se lê no Considerando (4)[66].

---

[62] V. tb. o art. 208.º-A CVM.

[63] Christoph Kumpan/Finn Schmidt, «Article 49 MiFID II», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, cit., p. 292--296, a p. 294.

[64] Seguimos de perto Emanuel de Fournoux et al., *A new framework for European financial markets*, LexisNexis, London, 2019, p. 83.

[65] Emanuel de Fournoux et al., *A new framework for European financial markets*, cit., p. 83.

[66] V. tb. o art. 208.º-A, 4, b), do CVM.

6. Deveres de comunicação, reporte e registo

A DMIF II pretende que as empresas de investimento que desenvolvam negociação algorítmica em Estado-Membro comuniquem às autoridades competentes a sua atividade, E isto quer no país de origem, quer no país da plataforma de negociação (art. 17.º, 2). A autoridade competente do Estado-Membro de origem pode exigir informação sobre a estratégia de negociação algorítmica, as suas atividades, e as medidas de *compliance* e de risco, bem como sobre os testes realizados aos seus sistemas[67].

A obrigação de reportar transações prevista no art. 26.º, 3, do RMIF também versa sobre os algoritmos: «Os reportes incluem, em especial, informações pormenorizadas relativas […] os algoritmos da empresa de investimento responsável pela decisão de investimento e pela execução da transação».

Para poder cumprir os deveres que a DMIF II lhe pretende impor, a empresa de investimento que aplique técnicas de negociação algorítmica de alta frequência está sujeita ao dever de efetuar registos precisos e cronológicos de todas as suas ordens, incluindo cancelamentos de ordens, ordens executadas e ofertas em plataforma de negociação», que podem ter de ser colocados à disposição da autoridade competente que o solicite (art. 17.º, 2, DMIF II)[68].

É também de registos que, no âmbito dos controlos pós-negociação, nos fala o art. 17.º do Regulamento Delegado 2017/589. Decorre do seu n.º 3 que as empresas de investimento devem manter «registos completos, exatos e coerentes das informações relativas à negociação e às contas». Devem também proceder à conciliação dos «registos eletrónicos das transações com as informações relativas às ordens e exposições em curso, tal como previsto pelas plataformas de negociação às quais enviam ordens, pelos seus corretores ou prestadores de DEA, pelos seus membros de compensação ou contrapartes centrais e pelos seus fornecedores de dados ou outros parceiros de negócios relevantes […]». O n.º 5 acrescenta que o controlo pós-negociação deve ser realizado pelos operadores responsáveis pelo algoritmo e pela função de controlo de risco da empresa de investimento.

---

[67] V. o art. 317.º-E, 3 a 5, do CVM.

[68] V. tb. os arts. 317.º-E, 6, e 317.º-F, 1, CVM.

Merece ainda referência o art. 5.º, 7, do Regulamento Delegado 2017/589, que obriga as empresas de investimento a manter registos de alterações substanciais ao *software* utilizado para a negociação algorítmica (quando, quem fez, quem aprovou, natureza da alteração).

7. Segurança e limites ao acesso

Estando a negociação algorítmica dependente da utilização da informática, as questões relacionadas com a segurança têm grande importância. O art. 18.º do Regulamento Delegado 2017/589 estabelece, no seu n.º 1, que as empresas de investimento devem adotar uma estratégia informática que esteja em conformidade com a sua estratégia empresarial e de risco e adpatada às suas atividades operacionais e riscos a que está exposta, que tenha por base uma organização informática fiável e que esteja em conformidade com um sistema eficaz de gestão de segurança informática. Além disso, as empresas de investimento devem «estabelecer e manter disposições adequadas de segurança física e eletrónica que minimizem os riscos de ataque contra os seus sistemas informáticos […] e que incluam uma gestão eficaz em termos de identificação e acesso» (n.º 2). Infrações significativas das medidas de segurança física e eletrónica devem ser dadas a conhecer à autoridade competente (n.º 3) e devem realizar-se anualmente testes de penetração e vulnerabilidade para simular ciberataques (n.º 4).

As empresas de investimento devem igualmente estar aptas a identificar todas as pessoas com direitos de acesso críticos aos seus sistemas informáticos e limitar número com esse acesso, bem como controlar o seu acesso aos sistemas informáticos para assegurar rastreabilidade (n.º 5).

8. O DEA

Vimos que o acesso eletrónico direto (*Direct Electronic Access* – DEA) envolve perigos significativos. A DMIF II, no seu art. 17.º, 5, ocupa-se do DEA. Se uma empresa de investimento proporciona esse acesso, «deve dispor de sistemas e controlos eficazes que assegurem a realização de uma avaliação e análise corretas da aptidão dos clientes que utilizam o serviço, que os clientes que utilizam o serviço estão impedidos de ultrapassar limiares de crédito e de negociação pré-estabelecidos e adequados, que a negociação por clientes que utilizam

o serviço é devidamente acompanhada e que os controlos de risco adequados impedem que a negociação seja suscetível de criar riscos para a própria empresa de investimento ou de criar ou contribuir para perturbações no mercado ou ser contrário ao disposto no Regulamento (UE) 596/2014 ou à regras da plataforma de negociação. É proibido o acesso eletrónico direto sem esses controlos». Além disso, uma empresa de investimento que proporciona o DEA «[…] é responsável por assegurar que os clientes que utilizem aquele serviço cumpram os requisitos da presente diretiva e as regras da plataforma de negociação. A empresa de investimento controla as transações a fim de identificar violações dessas regras, condições anormais de negociação ou comportamentos sucetíveis de envolver abuso de mercado e que devam ser comunicados à autoridade competente […]»[69].

9. Responsabilização

A identificação do que cada um deve fazer tem grande relevo para se poder encontrar quem deva ser responsabilizado pela prática de eventuais ilícitos. No Regulamento Delegado 2017/589 o art. 1.º logo determina que as empresas de investimento devem estabelecer e controlar os sistemas e algoritmos de negociação através de dispositivo de governo claro e formalizado que defina linhas claras de responsabilização. E estas devem dizer respeito também aos procedimentos para aprovação da conceçção, introdução e atualizações dos algoritmos de negociação e para resolução de problemas identificados quando os algoritmos de negociação são controlados. Exige-se ainda, designadamente, uma separação de tarefas e responsabilidades.

Tendo lugar a externalização e contratação de software ou hardware usado na negociação algorítmica, o art. 4.º, 1, daquele Regulamento afirma a responsabilidade plena das empresas de investimento pelas suas obrigações nele estabelecidas. Daí que seja tão importante garantir que as empresas de investimento tenham os conhecimentos necessários (art. 4.º, 2).

Como as empresas de investimento atuam através do seu pessoal, os arts. 2.º e 3.º do Regulamento Delegado 2017/589 dão atenção à preparação e competências daquele. Nomeadamente, quanto à forma

---

[69] V. tb. o art. 317.º-H do CVM.

como os sistemas de negociação algorítmica e os algoritmos de negociação da empresa de investimento funcionam.

10. Criadores de mercado

Vimos que a atuação dos criadores de mercado que recorrem à HFT envolve riscos acrescidos. Se, por um lado, contribuem para aumentar a liquidez no mercado, por outro, ao lidarem com grandes volumes, a forma como reagem pode ter consequências significativas.

O Regulamento Delegado (UE) 2017/578, de 13 de junho de 2016, olhou para os requisitos em matéria de acordos e sistemas de criação de mercado. Desde logo, e como se lê no Considerando (9), os criadores de mercado deverão «cumprir um conjunto mínimo de requisitos em termos de presença, volume e *spread*, em todos os casos». Acresce que os sistemas de criação de mercado das plataformas de negociação devem ter em conta a «efetiva contribuição dos participantes nos sistemas para a liquidez da plataforma». O regime em causa é particularmente importante porque é aplicável aos mercados regulamentados, aos sistemas de negociação multilateral e aos sistemas de negociação organizados (v. o Considerando (3))[70].

O Regulamento Delegado 2017/578 dá, com efeito, especial importância ao conteúdo do acordo que deve ser celebrado entre a empresa de investimento e a plataforma de negociação. Procura-se evitar a diminuição súbita de liquidez, especialmente em condições de tensão no mercado, como se comprova pela leitura do Considerando (8)[71].

11. Exigências relativas às plataformas. Os custos de cancelamento

Os riscos envolvidos na negociação algorítmica e, em particular, na HFT, foram também tratados pela DMIF II estabelecendo exigências relativamente às plataformas de negociação. O art. 48.º ocupa-se, designadamente, dos *circuit breakers* e dos sistemas de negociação

---

[70] V. tb. o art. 18.º, 5, da DMIF II e a remissão para os arts. 48.º e 49.º. Porém, há que ter em conta que as regras da DMIF II quanto aos criadores de mercado não abrangem toda a atividade destes: v. Emanuel de Fournoux et al., *A new framework for European financial markets*, cit., p. 92.

[71] Sobre a negociação algorítmica com estratégia de criação de mercado v. tb. o art. 317.º-G CVM.

eletrónica. Ali se estabelece que os mercados regulamentados devem ter sistemas, procedimentos e mecanismos eficazes para garantir que os sistemas de negociação «são resistentes, têm capacidade suficiente para lidarem com picos de ordens e mensagens, são capazes de assegurar a negociação ordenada em condições de forte tensão no mercado, estão plenamente testados para garantir o cumprimento dessas condições e são regidos por mecanismos de continuidade das atividades que asseguram a manutenção dos seus serviços, caso se verifique uma falha dos seus sistemas de negociação»[72].

A identificação de situações de tensão é importante para evitar um efeito de bola de neve e o agravar dos problemas. Daí que o art. 6.º, 2, do Regulamento Delegado 2017/578 obrigue as plataformas de negociação a «definir os parâmetros que identicam condições de tensão no mercado».

O sistema de *flagging* (sinalização) que ficou a constar do art. 48.º, 10, da DMIF II é, igualmente, merecedor de destaque. Resulta daquele preceito que os Estados-Membros devem exigir «que os mercados regulamentados sejam capazes de identificar, através de sinalização dos membros ou participantes, as ordens geradas por negociação algorítmica, os diferentes algoritmos utilizados para a criação das ordens e as pessoas pertinentes que dão essas ordens», sendo tais informações disponibilizadas às autoridades competentes a pedido das mesmas. Como se lê no Considerando (67) da DMIF II, essa sinalização permite, designadamente, que as autoridades competentes identifiquem as ordens provenientes de diferentes algoritmos e reconstruir e avaliar as estratégias utilizadas.

A sincronização dos relógios profissionais tornada necessária pelo art. 50.º, 1, da DMIF II quanto às plataformas de negociação e respetivos membros ou participantes tem grande importância para um rigoroso registo das transações[73], sendo essencial para as identificar e distinguir[74]. Como se lê no Considerando (1) do Regulamento

---

[72] V. tb. o art. 208.º-A CVM.

[73] Christoph Kumpan/Finn Schmidt, «Article 50 MiFID II», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, cit., p. 296--299, a p. 296.

[74] Johannes Karremans/Magnus Schoeller, «MiFID II between European rule-making and national market surveillance: the case of high-frequency trading», cit., a p. 42.

Delegado (UE) 2017/574, de 7 de junho de 2016, a sincronização dos relógios contribui «para assegurar que os dados de transparência pós-negociação podem fazer imediatamente parte de um sistema fiável de prestação de informações consolidadas. É igualmente essencial para realizar o acompanhamento das ordens pelas plataformas e para detetar casos de abuso de mercado, e permite uma comparação mais clara entre a transação e as condiçoes de mercado prevalecentes no momento da sua execução». De acordo com o art. 1.º do Regulamento Delegado 2017/574, a sincronização deve ser efetuada «com o tempo universal coordenado (UTC) emitido e mantido pelos centros de definição do tempo listados no último relatório anual sobre atividades de tempo do Bureau international des poids et mesures» e também «com o UTC difundido por um sistema de satélites, desde que qualquer desvio em relação ao UTC seja tido em conta e eliminado do carimbo temporal»[75]. A sincronização dará um precioso contributo para que as plataformas de negociação possam demonstrar que controlam, em tempo real, os elementos do seu sistema de negociação de acordo com o disposto no art. 13.º do Regulamento Delegado 2017/584, de 14 de julho de 2016[76].

A colocação de ordens que não se destinam a ser executadas é, como foi já dito, um comportamento que pode inserir-se numa estratégia de manipulação de mercado. Esse comportamento pode ser desincentivado através do que seja cobrado por tais cancelamentos. No Considerando 65 da DMIF II lê-se que é «também necessário garantir que as estruturas de comissões das plataformas de negociação […] não estejam organizadas de maneira a fomentar perturbações no mercado. Convém, por isso, que as plataformas de negociação estejam habilitadas a ajustar as comissões impostas às ordens canceladas em função do período de tempo em que a ordem foi mantida […]. Os Estados-Membros deverão também poder autorizar as plataformas de negociação a impor uma comissão mais elevada para a colocação de ordens que

---

[75] V. tb. as ESMA *Guidelines on transaction reporting, order record keeping and clock synchronisation under MiFID II*, 2017, ESMA/2016/1452, e ESMA *Final Report. Draft Regulatory and Implementing Technical Standards MiFID II/MiFIR*, ESMA/2015/1464.

[76] Lembrando isso mesmo, Christoph Kumpan/Finn Schmidt, «Article 50 MiFID II», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, cit., p. 297.

sejam posteriormente canceladas ou aos participantes que coloquem um elevado rácio de ordens canceladas e aos que põem em prática uma técnica de negociação algorítmica de alta frequência […]»[77].

12. Combater a fuga para os *dark markets*

Referimos que a volatilidade pode contribuir para levar investidores a afastarem-se dos mercados em que a HFT está presente, optando por alternativas menos transparentes. O art. 23.º, 1, do RMIF revela uma certa tendência para a concentração das transações sobre ações admitidas à negociação em mercado regulamentado ou negociadas em plataforma de negociação «num mercado regulamentado, MTF ou internalizador sistemático, ou numa plataforma de negociação considerada equivalente […]» (mas v. as exceções ali contidas)[78].

VII. PERSPETIVAS PARA O FUTURO

O atual enquadramento jurídico que a negociação algorítmica conhece na UE confere grande autonomia aos Estados-Membros em relação à supervisão dos mercados, o que cria um risco de fragmentação. E isto apesar de a ESMA garantir uma certa coordenação e de os Regulamentos Delegados reduzirem alguma margem de atuação dos Estados-Membros[79].

No final de 2020 a ESMA abriu uma consulta para receber contributos relativamente ao impacto das exigências da DMIF II e do RMIF

---

[77] V. tb. o art. 48.º, 9, da DMIF II e os arts. 223.º-A, 1, a), e 223.º-A, 2, c), do CVM.

[78] Peter Gomber/Ilya Gvozdevskiy, «14. Dark Trading under MiFID II», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and MiFIR*, cit., p. 386, consideram que dali resultará uma migração de «volumes from OTC trading into regulated venues». V. tb. o art. 201.º-B CVM. ». Quanto ao papel da DMIF I no sentido da abolição das «concentration rules», Guido Ferrarin/Paolo Saguato, «11. Governance and Organization of Trading Venues», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and* MiFIR, cit., p. 285-314, a p. 291, salientando a intenção de fragmentar os mercados e aumentar a concorrência e, com isso, a inovação e o avanço tecnológico. V. tb. os arts. 198.º, 5, e 201.º-B do CVM.

[79] V. Johannes Karremans/Magnus Schoeller, «MiFID II between European rule-making and national market surveillance: the case of high-frequency trading», cit., p. 40 e s., p. 46.

no que diz respeito à negociação algorítmica, incluindo a HFT. A consulta abrange o regime de autorização, as disposições sobre as empresas de investimento que usam algoritmos de negociação e recorrem à HFT e as normas sobre plataformas de negociação que admitem ou permitem a atuação dessas empresas. Cobre ainda, designadamente, os *circuit breakers*, os *speedbumps* e a transparência das transações. A consulta encerrou em 12 de março de 2021, devendo seguir-se um relatório da ESMA.

Na altura em que escrevemos estas linhas ignoramos se já foi elaborado esse relatório, mas talvez dele possam resultar propostas de alteração relevantes. É, porém, difícil adivinhar qual será o sentido das mesmas, pois as tensões existem. O governo holandês, durante a discussão que conduziu à DMIF II, resistiu muito à nova regulação porque na Holanda estava a sede de importantes *traders*, enquanto a Itália e a França tiveram posição mais aberta à intervenção[80].

O *Brexit* pode trazer problemas ou oportunidades no futuro[81]. O Reino Unido e a UE concluíram um Acordo de Comércio e Cooperação (*Trade and Cooperation Agreement*) em 24 de dezembro de 2020, mas este não se ocupa dos serviços financeiros, tendo sido aceite que o regime de terceiros Estados se aplicará e adotando-se um Regime de Permissão Temporária (*Temporary Permissions Regime*)[82]. Mas, no que diz respeito a terceiros Estados, a DMIF II deixa muito à atuação dos Estados-Membros[83].

Muitos *traders* e muitas plataformas queixaram-se de excessiva regulação e dos elevados custos relacionados com a armazenagem de in-

---

[80] Johannes Karremans/Magnus Schoeller, «MiFID II between European rule-making and national market surveillance: the case of high-frequency trading», cit., p. 33.

[81] V., p. ex., Jeffreu Golden, «The New European Capital Markets», in Cally Jordan (ed.), *International Capital Markets: Law and Institutions*, 2nd. ed., OUP, Oxford, 2021, p. 179-214, p. 213.

[82] George Walker, «Financial Markets and Exchanges», in Michael Blair/George Walker/Stuart Willey (ed.), *Financial Markets and Exchanges Law*, 3rd. ed., OUP, Oxford, 2021, p. 3-52, a p. 50.

[83] V. os arts. 39.º e ss. da DMIF II e, p. ex., Emmanuel de Fournoux et al., *A new framework for European Financial Markets,* cit., p. 204 e ss., e Nico Leslie/Aaron Taylor, «Markets in Financial Instrumentos Directive II (MiFID II)/Markets in Financial Instruments Regulation (MiFIR)», in Jonathan Herbst/Simon Lovegrove (ed.), *Brexit and Financial Regulation*, OUP, Oxford, 2020, p. 243-264, a p. 259 e ss..

formação e com o pessoal que têm de contratar. No entanto, a IA e os algoritmos permitiram também que surgisse a *Regulatory Technology* (RegTech)[84], tornando possível a redução de custos relacionados com o cumprimento de normas. Por outro lado, está aí também a *Supervisory Technology* (SupTech)[85], embora muito dependente dos orçamentos disponíveis no que diz respeito, desde logo, à capacidade de analisar a informação recebida e à sua segurança[86].

Perante os riscos associados à utilização de algoritmos e à HFT, alguma esperança tem sido depositada em sistemas que atrasam a introdução de ordens (*Speedbumps*), ocorrendo esta apenas após decurso de algum tempo depois de recebidas pelos mercados[87]. Trata-se de uma solução que tinha sido proposta pelo Parlamento Europeu, mas que não vingou[88]. Nos EUA, a *Investor Exchange* tem um *Speedbumb* de 350 microsegundos e soluções semelhantes estão presentes em alguns mercados europeus (*London Metal Exchange, Eurex*)[89]. Um *Speedbump* atrasa a visibilidade e, desse modo, afeta a capacidade de deteção que os algoritmos têm.

Veremos para que lado cai o pêndulo.

---

[84] Adrienne Héritier/Magnus Schoeller, «Governing finance in Europe: a centralisation of rule-making?», in Adrienne Héritier/Magnus Schoeller (ed.), *governing Finance in Europe*, cit., p. 19, definem a RegTech como «a digitally based in-time observation of financial market transactions monitoring their compliance with existin regulations». Por sua vez, Patrick Armstrong, «RegTech and SupTech – change for markets and authorities», in ESMA, *Report on Trends, Risks and Vulnerabilities*, n.º 1, 2019, p. 42 a 46, a p. 42, considera que se trata de «technology, particularly information technology, used in the context of regulatory compliance, including tasks such as risk management».

[85] SupTech designa a «technology used by supervisory authorities»: Patrick Armstrong, «RegTech and SupTech – change for markets and authorities», cit., p. 42.

[86] V., sobre a necessidade de os supervisores e reguladores se adaptarem a um «more intense, data-driven supervisory process», Patrick Armstrong, «RegTech and SupTech – change for markets and authorities», cit., p. 43. Como o autor também lembra, a resposta dos regulados pode ter em vista «to side-step regulations».

[87] V., desenvolvidamente, ESMA, *Consultation Paper. MiFID II/MiFIR review report on Algorithmic Trading*, 18 December 2020, ESMA70-156-2368, p. 82 e ss..

[88] Johannes Karremans/Magnus Schoeller, «MiFID II between European rule-making and national market surveillance: the case of high-frequency trading», cit., p. 39.

[89] ESMA, *Consultation Paper. MiFID II/MiFIR review report on Algorithmic Trading*, cit., p. 85.

REFERENCES

Almeida, Miguel Santos, «High-frequency trading – Regulação e compliance no contexto da nova DMIF II», in Paulo Câmara (coord.), *O novo direito dos valores mobiliários*, Almedina, Coimbra, 2017, p. 427-446.

Armstrong, Patrick, «RegTech and SupTech – change for markets and authorities», in ESMA, *Report on Trends, Risks and Vulnerabilities*, n.º 1, 2019, p. 42-46

Cognac, Pierre-Henri, «Algorithmic Trading and High-Frequency Trading (HFT)», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and MiFIR*, cit., p. 469-485

Ferrarin, Guido/Saguato, Paolo, «11. Governance and Organization of Trading Venues», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and* MiFIR, OUP, Oxford/New York, 2017, p. 285-314

Fournoux, Emanuel de, et al., *A new framework for European financial markets*, LexisNexis, London, 2019

Fox, Merritt, «MiFID II and equity trading. A US View», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and MiFIR*, OUP, Oxford/New York, 2017, p. 487-525

Golden, Jeffreu, «The New European Capital Markets», in Cally Jordan (ed.), *International Capital Markets: Law and Institutions*, 2nd. ed., OUP, Oxford, 2021, p. 179-214

Gomber, Peter/Gvozdevskiy, Ilya, «14. Dark Trading under MiFID II», in Danny Busch/Guido Ferrarini (ed.), *Regulation of the EU Financial Markets. MiFID II and MiFIR,* OUP, Oxford/New York, 2017, p. 363-389

Gonçalves, Telma Filipa Batista, «Estudo sobre os desafios da negociação algorítmica e de alta frequência na eficiência financiera e na integridade do mercado – novos desenvolvimentos regulatórios», *Direito dos Valores Mobiliários II*, ebook, IVM/AAFDL, 2018, p. 286-355

Héritier, Adrienne/Schoeller, Magnus, «Governing finance in Europe: a centralisation of rule-making?», in Adrienne Héritier/Magnus Schoeller (ed.), *Governing Finance in Europe*, Elgar, Cheltenham/Northampton, 2020, p. 1-30

Karremans, Johannes/Schoeller, Magnus, «MiFID II between European rule-making and national market surveillance: the case of high-frequency trading», in Adrienne Héritier/Magnus Schoeller (ed.), *Governing Finance in Europe*, Elgar, Cheltenham/Northampton, 2020, p. 32-51

Kroszner, Randall/Shiller, Robert, *Reforming U.S. Financial Markets*. Reflections Before and Beyond Dodd-Frank, MIT Press, Cambridge, Massachusetts/London, 2011

Kumpan, Christoph/Schmidt, Finn, «Article 49 MiFID II», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, Beck-Hart-Nomos, München/Oxford/Baden-Baden, 2019, p. 292-296, «Article 50 MiFID II», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, Beck-Hart-Nomos, München/Oxford/Baden-Baden, 2019, p. 296-299

Lehmann, Matthias, «Article 4 MiFID II», in Matthias Lehmann/Christoph Kumpan (ed.), *European Financial Services Law*, Beck-Hart-Nomos, München/Oxford/Baden-Baden, 2019, p. 22-34

Leslie, Nico/Taylor, Aaron, «Markets in Financial Instrumentos Directive II (MiFID II)/Markets in Financial Instruments Regulation (MiFIR)», in Jonathan Herbst/Simon Lovegrove (ed.), *Brexit and Financial Regulation*, OUP, Oxford, 2020, p. 243-264

Lin, Tom, «The New Market Manipulation», in 66 EmoryLJ, 2017, p. 1253-1314

McGowan, Michael, «The rise of computerized high frequency trading: use and controversy», *Duke Law & Technology Review*, 9, 2009, s/p.

Quelhas, José Manuel, «High-frequency trading (HFT)», *BCE*, LVIII, 2015, p. 369-399

Scopino, Gregory, *Algo Bots and the Law. Technology, Automation, and the Regulation of Futures and other Derivatives*, CUP, Cambridge/New York/Port Melbourne/New Delhi, 2020

Spindler, Gerald, «Controlo of Algorithms in Financial Markets. The Example of High-Frequency Trading», in Martin Ebers/Susana Navas (ed.), *Algorithms and Law*, CUP, Cambridge/New York/Port Melbourne/New Delhi, 2020, p. 207-220

Temporale, Ralph, *Europïsche* Finanzmarktregulierung, Schäffer-Peowschel, Stuttgart, 2015

Thelen, Martin Konstantin, *Dark Pools*, Duncker & Humblot, Berlin, 2019

Walker, George, «Financial Markets and Exchanges», in Michael Blair/George Walker/Stuart Willey (ed.), *Financial Markets and Exchanges Law*, 3rd. ed., OUP, Oxford, 2021, p. 3-52

3

The use of Big Data and Artificial Intelligence to prevent and detect fraud

(https://doi.org/10.47907/livro2021_4c4)

José Ricardo Marcondes Ramos[1]

Abstract:
The development and increased adoption of different IT trends is fostering the evolution and application of Big Data and Artificial Intelligence in many contexts of society and different business sectors. The financial services are the leading field in investment in advanced technology for continuous monitoring, transaction analysis, anomaly detection, data review, pattern recognition and other emerging techniques. In this article, we are going to analyse the role played by Big Data and artificial intelligence for fraud detection, analysing algorithms applied to prevent and detect payment frauds and the use of these techniques in the context of corporate fraud and investigations against financial statement frauds.

Keywords: big data; artificial intelligence; data mining; fraud detection; fraud prevention; forensic accounting.

[1] PhD student at the University of Coimbra. Univ. Coimbra, University of Coimbra Institute for Legal Research, Fac. Law - Collaborator.

I. INTRODUCTION

Ranging from the digital wallet to all sorts of smart gadgets oriented by sensors that perceive their surroundings[2] – such as autonomous vehicles, smart watches or fridges and everything that stands in between –, the development of new technologies is guiding the digital transformation of society, changing many aspects of life and introducing new ways of social interaction. Within the financial sector, for example, just as the development of ATMs (Automated Teller Machines) and online banking portals were revolutionary in the 1960s-1970s and 1990s, respectively, the creation of new technologies for financial transactions and the transfer of funds, such as crypto currencies and smart mobile payment systems (like Apple Pay, Samsung Pay, Google Pay and Amazon Pay, to name a few), are gradually replacing the need and the use of cash and facilitating the direct transfer of money and the purchase of goods and services[3].

As the rising use of electronic gadgets, due to this digital revolution, is increasing the production of digital information[4] (for instance, more than 98% of all information stored is currently electronic

---

[2] The ability to perceive the surrounding environment is usually associated with cloud-based systems that guide these gadgets with the use of the Internet of the Things (IoT), technology which "refers to the concept of connecting things like objects, people and animals, to the Internet using sensors that allow them to send and receive data in real-time". Bawack, R.E., Fosso Wamba, S. and Carillo, K.D.A. A framework for understanding artificial intelligence research: insights from practice. *Journal of Enterprise Information Management*, Vol. 34 No. 2, 2021, p. 653.

[3] Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Forensic Science International: Digital Investigation*, Volume 33, 2020, p. 01-02.

[4] As outlined by Bernard Marr "[w]e have created more data in the past two years than in the entire previous history of mankind. By 2020, it is predicted that about 1.7 megabytes of new data will be created every second, for every human being on the planet. This data is coming not just from the tens of millions of messages and emails we send each other every second via email, WhatsApp, Facebook, Twitter, etc. but also from the one trillion digital photos we take each year and the increasing amounts of video data we generate (every single minute we currently upload about 300 hours of new video to YouTube and we share almost three million videos on Facebook). On top of that, we have data from all the sensors we are now surrounded by". Marr, Bernard. Big Data in Practice. How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. Chichester: Wiley, 2016, p. 02.

whereas this figure was near 25% in 2000[5]), the improvement in the ability to collect and store data as well as to analyse different types of digital data is fostering transformations "from the way banks and shops operate to the way we treat cancer and protect our world from terrorism"[6]. From the two advertising giants Google and Facebook, which developed a business model of targeted advertising using Big Data and artificial intelligence techniques applied to massive databases of personal data gathered from its platforms; to the Royal Bank of Scotland[7], which applied data analytics techniques to redesign its customer relationship, creating a more personal service (the so-called "personology" philosophy) based on enormous amounts of information about its clients, there are plenty of examples of the usage of new informational technologies to boost efficiency for both business and government organizations[8].

Alongside the expansion of electronic databases, the development and increased adoption of different IT trends like the Internet of the Things (IoT), business intelligence and analytics (BI&A), Big Data, cloud computing and Machine Learning (ML) is fostering the evolution and application of yet another disruptive technology, namely, Artificial Intelligence (AI)[9] [10]. Based on its unique abilities to perceive

---

[5] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Managerial Auditing Journal*, Vol. 34 No. 3, 2019, p. 270.

[6] Marr, Bernard. Big Data in Practice. How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. *Cit.*, p. 01.

[7] As Bernard Marr describes, "RBS use data on their customers, including their account transactional history and personal information, to determine what products or services would be most useful" Marr, Bernard. Big Data in Practice. How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. *Cit.*, p. 84.

[8] As noted by Martin Fleming, VP and Chief Economist at IBM, "AI technology has the potential to increase the productivity of workers as well as productivity in all walks of life". Wilson, C. (2019), "IBM Tech trends to watch in 2020 . . . and beyond", IBM.

[9] Bawack, R.E., Fosso Wamba, S. and Carillo, K.D.A. A framework for understanding artificial intelligence research: insights from practice. *Cit.,* p. 645 and 655-657.

Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies.. *Cit.,* p. 01-02.

[10] There are several different perspectives to understand artificial intelligence, such as a *field of study* in which AI is perceived "as the branch of knowledge that investigates the possibility of giving human intelligence capabilities to nonhuman entities", a *concept* that understands it as "an abstract concept that corresponds to any manifestation of human intelligence by machines or technology", an *ability* by which

the environment, learn from experience, understand intention and context and take appropriate action with autonomous decisions[11] – carried out using complex algorithms that recognize patterns, understand written and spoken words, identify images and make predictions and recommendations – AI is one of the main technological and strategic trends in the digital transformation of business and society today, being the main responsible for business automation processes[12].

Although its consolidated adoption is so far limited to a handful of industries such as financial services, healthcare, marketing and fraud detection, there are an increasing number of studies and experiments designed to test the application of artificial intelligence in fields like education, telecommunication, transportation, automotive, energy and so on[13]. In this article, we will analyse the use of artificial intelligence in the combat of financial fraud, with a focus on algorithms used to prevent the occurrence of payment frauds and money laundering as well as to detect financial statement fraud in the context of forensic accounting. In order to do so, firstly, we are going to analyse the role played by digital tools, Big Data and artificial intelligence in the context of forensic investigations. Thus, our focus will be moved for fraud detection algorithms applied to prevent and detect payment frauds, notably, credit card frauds and, finally, the use of these techniques in the context of corporate fraud and investigations.

II. FROM FORENSIC SCIENCE AND DIGITAL FORENSICS TO ARTIFICIAL INTELLIGENCE AND BIG DATA APPLIED TO FRAUD DETECTION AND PREVENTION

Just as new technologies are facilitating and transforming several aspects of life interaction, the emergence of new technological trends

---

"AI is a skill given to a technology artifact for it to behave like an intelligent human being" and even as a *system*, perspective that sees AI as a set of technologies that "can perceive, learn, reason, assist in decision-making and solve problems in ways like humans". Bawack, R.E., Fosso Wamba, S. and Carillo, K.D.A. A framework for understanding artificial intelligence research: insights from practice. *Cit.,* p. 651-652.

[11] Despite the different points of view to understand artificial intelligence, the four capabilities of perception, comprehension, learning and acting are seen by practitioners as the main features that characterize Artificial Intelligence. Idem, p. 651.

[12] Idem, p. 652.

[13] Idem, p. 658.

is also being used to leverage both old and new criminal behaviours[14]. Either by tampering or meddling with technical and technological mechanisms (such as hardware, software, network protocols and cryptography) or by smoothing social engineering schemes in order to exploit personal weaknesses and enable fraud[15], the development of new operational techniques based on technological advances and emerging gadgets is aiding the occurrence of different forms of fraud, money laundering and other underground criminal activity[16].

If, however, technology may help crime be committed, it is also evolving to improve old investigative methods as well as to develop new ones[17]: first, the growing essentiality of digital gadgets (be it computers, smartphones, tablets or others) has led to the development of a whole new field of digital forensics[18] specialized in the extraction

---

[14] As Richard Bolton and David Hand argue "in recent years, the development of new technologies (which have made it easier for us to communicate and helped increase our spending power) has also provided yet further ways in which criminals may commit fraud. Traditional forms of fraudulent behavior such as money laundering have become easier to perpetrate and have been joined by new kinds of fraud such as mobile telecommunications fraud and computer intrusion". Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review, *Statistical Science*, Vol. 17, No. 3 (Aug., 2002), Institute of Mathematical Statistics, p. 235.

[15] Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Journal of Intellectual Capital*, 2020, Vol. 21 No. 2, pp. 259. Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Cit.,* p. 06-07. van Beek, H.M.A.; van den Bos, J.; Boztas, A.; van Eijk, E. J.; Schramp, R.; Ugen, M. Digital forensics as a service: Stepping up the game. *Forensic Science International: Digital Investigation*, Volume 35, 2020, p. 01.

[16] As Sunger Gee points out, "[n]ew technology allows for new, more convenient payment methods for consumers and also provides new opportunities for money laundering ... [In addition,] criminals can mix the old with the new to move money to further reduce the risks of detection". Gee, Sunder. Fraud and fraud detection: a data analytics approach. New Jersey: Ed. Wiley, 2015, p. 257.

[17] Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology. *Accounting Research Journal*, 28(1), p. 06.

[18] As a branch of forensic science, digital forensics is responsible for the process of identification, collection, processing and interpretation of digital data from any given device and, as such, can be understood as "the process of applying scientific methods to analyze stored information and to determine the events of a particular incident, thus making evidence usable in court". Oliveira Júnior, Edson; Zorzo, Avelino F.; Neu, Charles Varlei (2020). Towards a conceptual model for promoting digital forensics experiments. *Forensic Science International: Digital Investigation*, 35(), 301014, p. 01. It is important to recognize, though, that digital investigation's

and analysis of data produced, stored and processed within these devices[19] – be it software, hardware or a combination of both[20]. Second, the improvement of data extraction and analysis capacity, enabled by new technological advances such as data mining and data analysis techniques, is helping to upgrade crime-related forensic methods in diverse areas such as neurocriminology, handwriting analysis and forensic accounting[21].

While the increasing adoption of new information and communication technology by society is fostering new investigative sources and techniques for traditional crimes[22], the enormous amount of digital data produced by new digital devices is fuelling the development of digital forensics as a new and independent field[23] specialized in the

---

appliance is not restricted to judicial controversies, also being commonly used in the corporate ecosystem as a preventive and investigative tool related to behavioral and disciplinary concerns. However, even recognizing that the digital forensics has several applications within the legal frameworks – namely, public sector security and operation as well as corporate investigations – it is essential to keep in mind that the "main purpose of digital evidence is to support or rebut a thesis or argument on which court decision is based on". V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *2020 43rd International Convention on Information, Communication and Electronic Technology* (MIPRO), 2020, p. 2096.

[19] van Baar, R.B.; van Beek, H.M.A.; van Eijk, E.J. (2014). Digital Forensics as a Service: A game changer. *Digital Investigation*, 2014, 11, p. 54.

[20] V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *Cit.,* p. 2095. Wu, Tina; Breitinger, Frank; O'Shaughnessy, Stephen. Digital forensic tools: Recent advances and enhancing the status quo. *Forensic Science International: Digital Investigation*, 2020, 34, p. 04. Netherlands Register of Court Experts NRGD, 2016. Standards 008.0 Digital Forensics. Technical Report Netherlands Register of Court Experts, p. 06.

[21] Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology. *Cit.*, p. 07-08.

[22] Illustrative examples are the use of Google searches and other online activities to prove premeditation and, in a more concrete stance, the extraction of information from the Apple Watch app to unveil the disappearance and assassination of Saudi dissident Jamal Khashoggi in Turkey. Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Cit.*, p. 260. Lee, Jae-Ung; Soh, Woo-Young. Comparative analysis on integrated digital forensic tools for digital forensic investigation. *IOP Conference Series: Materials Science and Engineering*, 2020, 834, 012034, p. 01.

[23] As Ferguson *et al* explain "the field of digital forensics, though relatively young, has earned the right to call itself a discipline, and that law enforcement and

understanding of how this data is produced and how it can be collected and analysed[24]. Usually, crimes and felonies involving different sorts of technologies are very technical in nature[25], which implies different types of analysis of hardware, software systems, malware, network protocols, APIs and cryptography[26]. Even though there are different criteria for classifying digital forensics tools, the diversity of data sources and the need for expertise on the underlying technology is the base for its taxonomy, which separates the digital forensics in different sub-fields such as computer forensics, software forensics, multimedia forensics, device forensics, network forensics, malware forensics and memory forensics[27].

Regarding the old investigative methodology, on the other hand, ever since the introduction of the fingerprinting method (the first

---

educational institutions are developing training to ensure that effective investigations can indeed be carried out in the digital world to support law enforcement". Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Cit.,* p. 260.

[24] van Beek, H.M.A.; van den Bos, J.; Boztas, A.; van Eijk, E. J.; Schramp, R.; Ugen, M. Digital forensics as a service: Stepping up the game. *Cit.,* p. 01. van Baar, R.B.; van Beek, H.M.A.; van Eijk, E.J. (2014). Digital Forensics as a Service: A game changer. *Digital Investigation*, 2014, 11, p. 54.

[25] As Bruce Nikkel explains, there are several kinds of different criminal activities exploiting different technological bases. Considering solely financial frauds, there are felonies raging from phishing, attacks against ATMs and payment card terminals, online banking trojans, rogue mobile banking apps, extortion and ransom attacks, online social engineering attacks, online money laundering and others. Once the crime could be committed throughout different technological means, the investigative process may vary. A good example is the practice of phishing, tricking people into giving personal or financial information through "spoofed" messages, which could be committed by SMS (smishing), voice (vishing), twitter (twishing). Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Cit.,* p. 03-05.

[26] Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Cit.*, p. 06.

[27] This taxonomy is proposed by Tina Wu *et al* (2020) as an updated version of the distinction made by the Netherlands Register of Court Experts (NRGD, 2016. Standards 008.0 Digital Forensics. Technical Report Netherlands Register of Court Experts, p. 08). As the authors describe, their version has two central differences: firstly, "due to the lack of available database forensic tools", the data base sub-field is placed under the software category; and, secondly, the taxonomy was extended to include the categories of malware and memory forensics. Wu, Tina; Breitinger, Frank; O'Shaughnessy, Stephen (2020). Digital forensic tools: Recent advances and enhancing the status quo. *Cit.,* p. 04.

significant investigation technique), forensic science[28] developed a series of scientific processes in fields ranging from biology, chemistry and physics to anthropology and accounting to assist investigators identify and enquire information and objects related to a crime scene (a mute witness of the crime) and collect relevant evidence such as stains, hair or DNA samples, soil and so on[29]. Although several of the forensic investigation techniques and tools are not particularly new – for instance, the creation of the polygraph machine, which set the standard for lie-detection methodology, dates back to the 1880s[30] – it was not until the end of the 20th century that the use of computers to perform investigative tasks contributed to the development of the digital forensics field[31]. Despite being used to perform, with enhanced capacity, traditional forensic tasks like fingerprinting, hair or DNA analysis and even autopsies, the development of the digital forensics field was highly influenced both by the ubiquitous adoption of new digital devices and the rising incidence of cybercrime[32].

As a scientific field, the development of new applied research[33] and new technologies helped forensics science to improve its methods

---

[28] The field of forensic science emerged and was developed due to the difficulty of unveiling the circumstances in which a crime may have occurred and to overcome the excessive reliance on confessions or witness testimony to identify the offender. With this goal, the field of forensic science developed a series of techniques and methods to aid the investigative process by acquiring, analysing and interpreting evidence through a coordinated process in order to base scientifically investigative conclusions. V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *Cit.*, p. 2094.

[29] V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *Cit.*, p. 2094.

[30] Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology. *Cit.,* p. 07.

[31] Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Cit.*, p. 259. V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *Cit.*, p. 2094-2095.

[32] Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Cit.*, p. 259.

[33] As Tina Wu *et al* describe, "[c]ompared to other domains, the digital forensics community has a very applied focus, meaning that we are not solving problems in theory but practically. Consequently, research endeavors frequently come with prototype implementations". Wu, Tina; Breitinger, Frank; O'Shaughnessy, Stephen. Digital forensic tools: Recent advances and enhancing the status quo. *Cit.,* p. 04.

and create advances in its techniques for interviewing and interrogation, handwriting analysis, data analysis and others[34]. With respect to the interviewing and interrogation processes, for example, lie-detection techniques previously based on alterations in breathing, blood pressure, pulse rate and sweat, measured by the polygraph, can now be performed by neurocriminology instruments, which identify whether someone is lying or telling the truth based on neural mapping and the areas of the brain displayed as active when the person is confronted with evidences of the crime[35]. Similarly, handwriting analysis, which used to be done personally by experts, can now be performed by algorithms that examine features such as pen pressure and letter dimensions[36].

Finally, the development of new technological advances is also revolutionizing audit and forensic accounting practices and, as a consequence, not only innovative techniques, such as word mapping software that identify bribery-related terms, are being used to fight corruption[37], but the emergence of new data processing capacities and statistical tools for fraud detection is also improving the fight against financial crimes, corporate fraud and money laundering[38] – topics that will be analysed next. As described by Michael Young[39],

> Forensic computers can be deployed to look for fraud. Based on years of accumulated experience, savvy forensic accountants at the big accounting and consulting firms have developed computerized searching tools that, once plugged into a company's general ledger system, will at high speed start combing through thousands of entries and kicking out those that for any number of reasons look unusual or suspicious.

---

[34] Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology.*Cit.,* p. 07.

[35] Idem, p. 07-08. The author also adds that "[u]sing similar technology, a recent study of New Mexico inmates using brain scans correctly predicted which prisoners were more likely to commit another crime once released".

[36] Idem, p. 07.

[37] Idem, p. 08.

[38] As Richard Bolton *et al* describes, "[p]rocessing these data sets in a search for fraudulent transactions or calls requires more than mere novelty of statistical model, and also needs fast and efficient algorithms: data mining techniques are relevant". Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 236.

[39] Young, Michael R. *Financial Fraud Prevention and Detection. Governance and Effective Practices*. New Jersey: Wiley, 2014, p. 169.

III. THE APPLICATION OF BIG DATA AND ARTIFICIAL INTELLIGENCE TO DETECT AND PREVENT FINANCIAL FRAUD

The expansion of digital data and the increasing importance of electronic databases for economic development and strategic management are reshaping many aspects of society as well as several different business sectors – so much so, that the market of Big Data has grown dramatically from U$16,1 billion in 2014 to more than U$50 billion by the end of 2016[40]. Despite its rising adoption in a growing number of industries, though, when it comes to the fight against fraud, the financial services sector is the leading field in investment in advanced technology for continuous monitoring, transaction analysis, anomaly detection, data review, pattern recognition and other emerging techniques[41].

This scenario is justified by several reasons such as the better cost and operational efficiency that data-driven or statistically based fraud-detection methodologies present, or its greater precision and improved detection power when compared to the classic human-driven approach[42]. However, aside from the fact that data analysis allows the entire database of financial transactions or journal entries to be tested instead of a selected sample[43], perhaps the main reason may be the essentiality of Big Data and analytics techniques when dealing with huge amounts of dynamic and constantly evolving data[44], after all "by processing massive volumes of information, fraud patterns may be uncovered that are not sufficiently apparent to the human eye"[45].

---

[40] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Cit.,* p. 270.

[41] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Managerial Auditing Journal*, Vol. 34 No. 5, 2019, p. 602-622.

[42] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. A Guide to Data Science for Fraud Detection. New Jersey: Wiley, 2015, p. 17-18.

[43] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 67.

[44] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 236.

[45] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. A Guide to Data Science for Fraud Detection. *Cit.*, p. 17.

As a rule, the use of these techniques is separated into two different stages: *data mining*, when the computer uses artificial intelligence, neural network techniques and advanced statistical tools (e.g. cluster analysis) to perform searches on large amounts of data with the specific goal of finding trends, patterns and relationships within the dataset, but without testing any pre-established hypothesis; and, *data analysis*, when the evaluation of each component of the data has the purpose of testing a hypothesis to be confirmed or dismissed, reaching a conclusion based on the inference from the findings[46]. The data analytics process may be subsequently separated into three other steps of *exploratory data analysis* (EDA) in which very little in known about the data's relationships where "hypotheses are formed and new patterns of features of the data are discovered"; *confirmatory data analysis* (CDA), when "testing takes place and the hypotheses are proven correct or false"; and *qualitative data analysis* (QDA), stage "used to draw conclusions from non quantitative or non-numerical data such as images or text"[47].

The hypotheses tested by the data analysis techniques, in turn, are developed using two types of methods, namely, *supervised* ones, in which selected samples of behaviours labelled as both fraudulent and non-fraudulent are applied to train algorithms that assign a suspicion score to evaluated cases; and *unsupervised methods*[48], in which the algorithm is trained with a baseline of what represents the normal behaviour and focus on detecting anomalies outlining observations that depart from this norm[49]. Once supervised methods learn with historical observation from which they extract patterns of fraudulent and non-fraudulent behaviour, it is mostly applied when there is a wide and complete dataset about each type of fraud[50]. This, however, requires not only that the training set be composed of both classes of cases, but also that the labelling of each is reliable and balanced,

---

[46] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 10-11.

[47] Idem.

[48] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 19-20.

[49] Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50, (3), 2011, p. 602.

[50] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review.*Cit.*, p. 237.

avoiding a biased algorithm caused either by over- or under-sampling and, thus, reducing the occurrence of false negatives and positives[51].

Furthermore, besides the lack of testable and open access data[52] (considering not all victims of fraud publicly disclosure this type of information) and the limited exchange of information regarding fraud detection methods (notably because "it does not make sense to describe fraud detection techniques in great detail in the public domain, as this gives criminals the information that they require to evade detection"[53]), the use of supervised methods faces yet another problem: the dynamic character of fraud, which makes it unable to detect new types of fraudulent behaviour or frauds that uses unknown mechanisms or methods[54]. As fraud detection algorithms evolve, so do fraudsters, adapting their approaches and strategies with inventive and refined new methods to make the fraudulent behaviour less apparent and detectable by upgraded algorithms[55]. Also, because there are still new criminals trying old methods (sometimes unaware of the consolidated detection techniques), the latest upgrades on the algorithms should be applied jointly with earlier tools[56].

More than showing that the growing availability of fraud data is an important driver for the improvement of fraud prevention and detection techniques, this dynamic character of fraud and the need to reduce the pernicious consequences of new frauds evidence the importance of regularly updating the algorithms throughout the fraud cycle and its four steps of fraud detection, investigation, confirmation and

---

[51] Beasens, Bart. Analytics in a Big Data World. The Essential Guide to Data Science and its Applications. New Jersey: Wiley, 2014, p. 165-166.

[52] Shiguo Wang goes further and argues that "[t]here are two kinds of critical suggestions concerning applying data mining technology on detecting fraud. One is lack of testable, open accessible data. The other is lack of mature methods and technologies". Wang, Shiguo. A comprehensive survey of data mining-based accounting-fraud detection research. *2010 International Conference on Intelligent Computation Technology and Automation*, ICICTA 2010, 1, 50.

[53] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 236.

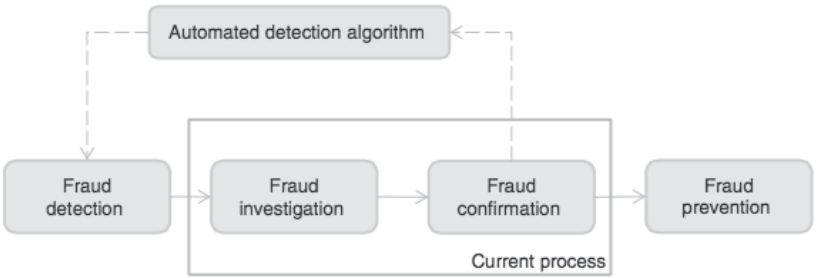
[54] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 237.

[55] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 18.

[56] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 236.

prevention (as shown in Figure 1)[57]. Although the required frequency for retraining the algorithm is influenced by several factors such as the volatility of the unknown fraud behaviour, the rate at which new cases are confirmed, the detection power of the existing model and the cost-benefit relation associated with the upgrading process, the main element for this feedback loop is the correct and reliable labelling of cases as fraudulent in a careful *ex post* analysis[58]. Besides, these limitations on the supervised methods also "illustrates the complementarity of supervised and unsupervised methods and motivates the use of both types of methods as complementary tools in developing a powerful fraud-detection and prevention system"[59].

Figure 1. The fraud cycle



Source: Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. **Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques.** *Cit.,* p. 23.

Alongside the supervised and unsupervised methods, there is a third complementary tool used to enhance the abilities for fraud detection, known as social network analysis. Based on researches showing that fraudsters seldom act in an isolated fashion and are usually highly interconnected with other fraudulent individuals and companies[60], this method creates a so-called spider construction (Figure 2) in order to analyse network-related information and identify potentially suspicious activities[61].

---

[57] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 11-12.

[58] Idem, p. 23.

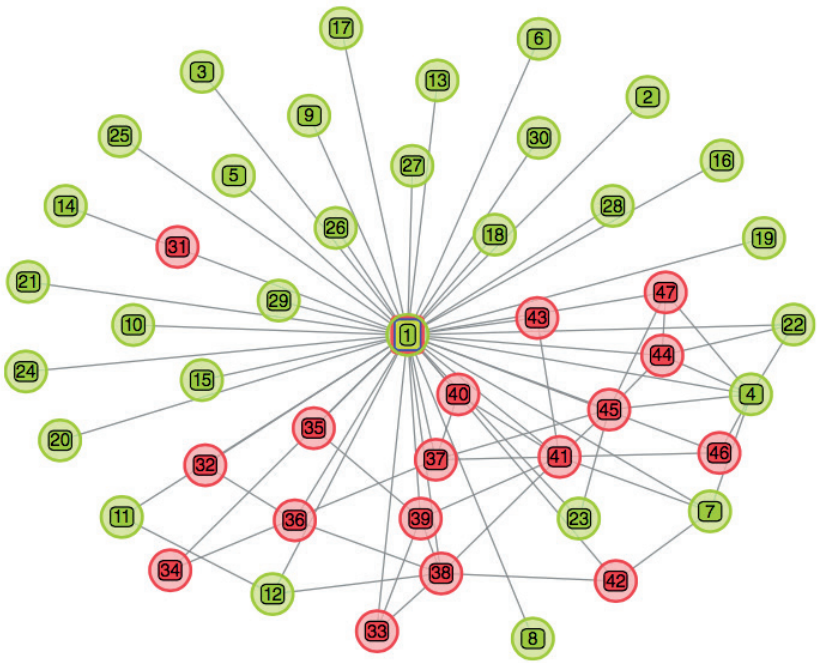
[59] Idem, p. 21.

[60] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 237-238.

[61] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Cit.,* p. 271. Beasens, Bart. Analytics in a Big Data World. The Essential Guide to Data Science and its Applications. *Cit.,* p. 166-167.

For instance, within the analysis of banking accounts, by examining variables such as the *fraudulent degree*, obtained by analysing the number of immediate contacts a node has and its number of direct fraudulent connections; the *triangles* it belongs to (that is, structure of three nodes connected to each other) and the potential of being a fraudulent node by integrating with a fraudulent triangle (after all "nodes that are involved in many suspicious triangles have a higher probability to commit fraud themselves"[62]); and the *cliques*, or extensions of triangles that may indicate undirected connections with other fraudsters, are important techniques used to identify money laundering and complex structure for carousel fraud[63].

Figure 2. Spider constructions of social network analysis



Green nodes represent legitimate individuals, while red ones represent fraud.

Source: Beasens, Bart. Analytics in a Big Data World. The Essential Guide to Data Science and its Applications. *Cit.,* p. 168.

---

[62] Idem, p. 168.

[63] Idem, p. 167-168. Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 04.

Even recognising that the identification and analysis of the agents involved in a specific group of financial transactions play a key role in the fight against frauds and money laundering, it is important to mention that the connection with other parties and potential fraudsters is not the only type of linked analysis performed by fraud detection algorithms. As such, there are also several levels of analysis that could be made (sometimes simultaneously) to reconstruct patterns of transactions and distinguish legitimate sets from illegitimate ones, including the participants engaged ("an obvious and simplistic illustration is the fact that a transaction with a known criminal may rouse suspicion"[64]), the individual transaction or its association with other sets of transactions ("a single deposit of just under $10,000 is not suspicious, but multiple such deposits are; a large sum being deposited is not suspicious, but a large sum being deposited and instantly withdrawn is"[65]) and even the geographical location of either the origination or destination of the funds, which orients rules such as "flag transactions from countries X and Y", based on international lists of countries considered to be at high risk of money laundering or that have some form of connection with terrorism[66].

It is important to mention, though, that these three fraud detection techniques are not mutually exclusive and, once each one focuses on a different aspect of fraud, are usually complementary. As a matter of fact, because each of these methods has different capacities and limitations, the development of an effective mechanism to prevent and detect different types of financial fraud is commonly based on the combination of the three, reinforcing each other's strength and compensating for its vulnerabilities[67]. Notably, however, the selection

---

[64] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 241.

[65] Idem.

[66] The evaluation of jurisdictions and their compliance to international standards related to the combat of money laundering and terrorist financing (AML/CFT) is made by the Financial Action Task Force (FAFT), which releases, three times a year, two lists containing the countries, first, with weak AML/CFT regimes and, second, under increased monitoring, that is, countries working with the FAFT to address their strategic deficiencies. Both lists can be accessed at https://www.fatf-gafi.org/publications/high-risk-and-other-monitored-jurisdictions/?hf=10&b=0&s=desc(fatf_releasedate).

[67] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 22.

of the ideal fraud prevention and detection system and the development of the most adequate algorithm depends on the type of crime to be avoided. In general, the solutions currently existing in the market aim at two different types of misconduct associated, on the one hand, with illegal financial flows in crimes such as credit card fraud, financial identity scams, money laundering and terrorist financing and, on the other hand, with different forms of financial fraud, which range from internal corporate crimes, such as money embezzlement and reckless management, to financial statement fraud in the form of market manipulation and investment fraud. Once each form of fraud assumes a different *modus operandi*, the recommended algorithm to detect and prevent it vary based on expertise and technique required, a topic that we shall address next.

IV. THE USE OF ARTIFICIAL INTELLIGENCE TO FIGHT PAYMENT FRAUDS, MONEY LAUNDERING AND TERRORIST FINANCING

Among other serious problems emerging in the financial sector, two main issues with major harmful consequences for the economy are located within the capital flows system: firstly, the occurrence of payment frauds, in the form of credit card fraud, account takeover and other scams; and the incidence of illegal financial flows, namely money laundering and terrorist financing. Regarding payment frauds, for example, in its latest report the European Central Bank points out that worldwide €1.80 billion was lost in 2018 alone, with a total of €0.94 billion defrauded from credit cards issued in the euro area[68]. As a fraudulent behaviour, credit card fraud integrates a larger form of financial crime known as *financial identity scams* in which a fraudster reaches for personal information of a victim in order to perform fraudulent money transfers or payments[69].

---

[68] European Central Bank. Sixth report on card fraud. August 2020, p. 02.

[69] As Arjan Reurink explains, there are several terms used to describe financial identity scams such as "identity crime", "identity theft", "identity fraud", "credit card fraud" and "payment fraud". Still according to the author, among this fraudulent behaviour it is possible to distinguish three main types of fraud, namely "financial identity theft, which entails the use of personal identifying information to establish credit lines in the name of the victim; criminal identity theft, which involves a criminal giving

With a similar *modus operandi* of, first, obtaining the client's identifying information – either using *technical subterfuge schemes* by installing malware and malicious software in digital devices (technique known as *pharming*[70]) or *social engineering schemes* for deceiving the victims into giving their information away (such as *phishing*[71]) or allowing someone to obtain it (using techniques like "skimming" or "shoulder surfing"[72]) – and, second, using it to realize financial gain, credit card fraud and account takeover alike can be classified essentially in two types: application and behavioural fraud[73]. While application

another person's identifying information to law enforcement; and identity cloning, whereby imposters, illegal immigrants, or wanted felons use the victim's personal information to establish a new life". Reurink, Arjan. Financial fraud: A literature review. MPIfG Discussion Paper, No. 16/5, Max Planck Institute for the Study of Societies, Cologne, 2016, p. 47.

[70] Arjan Reurink explains that technical subterfuge schemes "are more technical in nature and rely much less on persuasion to entice victims into the scheme, which enables a much wider victim base". Regarding the pharming technique, for example, "fraudsters send out e-mails which, when opened, plant malware – malicious software – in the victims' personal computers. The malware then directs traffic from those PCs that is destined for a legitimate website, say, a bank, to the pharmer's bogus website, which looks just like the real one. Without the victim's knowledge or consent, all the information the victim thinks is being sent to the bank's website is sent directly to the pharmer. Another possible mode of operation for pharmers is to alter a website's internet protocol (IP) address in the domain name server (DNS). In so doing, pharmers redirect all users who type in the URL (the web address) of, say, a bank to the illegitimate website controlled by the pharmer". Idem, p. 48.

[71] "In a typical phishing attack, a scam artist pretending to be an agent from a bank or credit card company sends out e-mails to customers in which the operator prompts them to click on a hyperlink that brings them to a website, controlled by the phisher, where they will be asked to further process their account details. To appear credible and to trick the recipient's into participating in the scheme, the scam artist's e-mails contain company logos and use scare tactics – such as threats of account closure – and urgency cues that short-circuit victims' elaboration on clues that could reveal the deceptive nature of the invitation". Idem, p. 48.

[72] According to Richard Bolton and David Hand "skimming" is a technique "where employees illegally copy the magnetic strip on a credit card by swiping it through a small handheld card reader", while "shoulder surfers" are fraudsters "who enter card details into a mobile phone while standing behind a purchaser in a queue". Alongside there fraudulent techniques, the authors also mention yet another in which "people posing as credit card company employees taking details of credit card transactions from companies over the phone". Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 238-239.

[73] Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Cit.,* p. 603.

fraud occurs when a fraudster uses false information or other people's information to obtain a credit card, behaviour fraud is characterized by the fraudulent use of other people's credit card or another payment means without its knowledge and approval. This last type of fraud, is separated into four types: mail theft, in which credit cards are intercepted before reaching the cardholder; stolen or lost card; counterfeit card and "card holder not present" fraud[74]. Similarly, frauds involving bank accounts range from the *account takeovers*, in which a fraudster takes control over an existing account and extracts its balance (sometimes even creating additional accounts and using the victims' credit lines); to *fictitious identity fraud*, where pieces of real information are combined to fabricate a fake identity in order to defraud the banking institution by establishing credit lines and stealing the money[75].

In the literature, there is widespread agreement that these types of payment fraud have several harmful consequences that affect directly and indirectly three groups of victims consisting of the *consumers* and *businesses* that had their financial identity stolen and credit cards/bank accounts misused; *merchants* and *credit providers* tricked into giving credit, money and goods to scammers; and, finally, *banks*, *credit card companies* and *e-retailers* whose brands were associated with these felonies and may need to review and reform their cyber security programmes and policies[76]. Due to its pernicious consequences, thus, it is in all stakeholder's interests to avoid the occurrence of payment fraud or, at least, to detect it as soon as possible in order to reduce direct losses and preserve the confidence in the whole payment system[77].

Because a typical credit card transaction has abundant data availability by recording hundreds of characteristics that describe each transaction in detail, credit card companies are not only among the early adopters of Big Data approaches, but also of the use of artificial intelligence algorithms, one of the main weapons at their disposal[78].

---

[74] Idem, p. 603. Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 238-239.

[75] Reurink, Arjan. Financial fraud: A literature review. *Cit.*, p. 48-49.

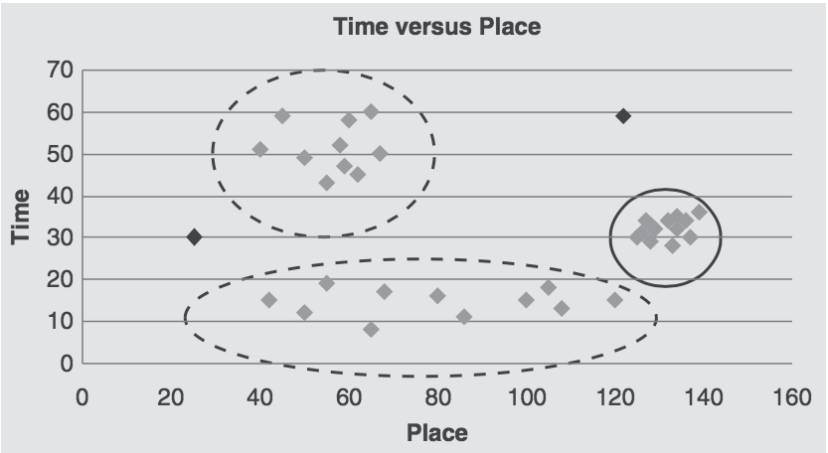
[76] Idem, p. 50.

[77] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 238.

[78] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 24.

By applying different data mining techniques like neural networks, random forests or logistic regression, the predictive models trained to detect credit card fraud analyses a broad variety of information attributed either to the transaction itself or the parties engaged in it. This includes both numerical attributes, like its amount, and categorical attributes, as the example of the merchant code and name, the date of the transaction and its type or its geographical location, among others[79]. Using these pieces of information, the algorithms use descriptive analytical methods such as *outlier detection techniques* in order to identify abnormal or anomalous behaviours that might indicate suspicious activities[80]. As Figures 3.1 and 3.2 show, by pinpointing transactions that deviate from the clusters of regular and frequently occurring pattern, it is possible to detect outliers that do not comply with the overall behaviour and, hence, flag a transaction for further human investigation[81].

Figure 3.1. Outlier detection at the data item level

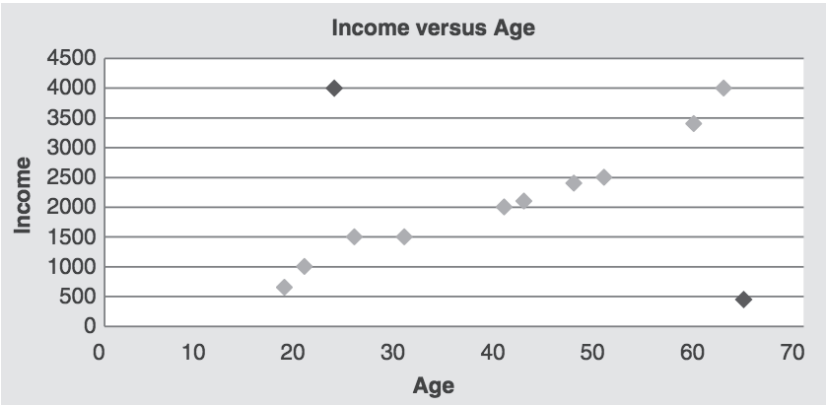


[79] Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Cit.,* p. 603-604.

[80] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 24.

[81] Idem, p. 24-25. Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 617.

Figure 3.2. Outlier detection at the data set level



Source: Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. Cit.,* p. 25.

The standards of normality used as benchmark to analyse transactions, in turn, are obtained by the application of a wide range of statistical tools, machine learning methods and data mining techniques which examine and identify both individual patterns of previous usage as well as general patterns of use and consumption according to the type of establishment, personal profile, age, location, etc.[82] (Figures 4.1 and 4.2). In addition, these predictive analytics tools also use algorithms trained to identify transaction patterns known to be intrinsically suspicious as the example of small purchases followed by big ones[83], a large number of online transactions made in a short period of time, the immediate use of a new card in a wide range of different locations as quickly as possible[84],

---

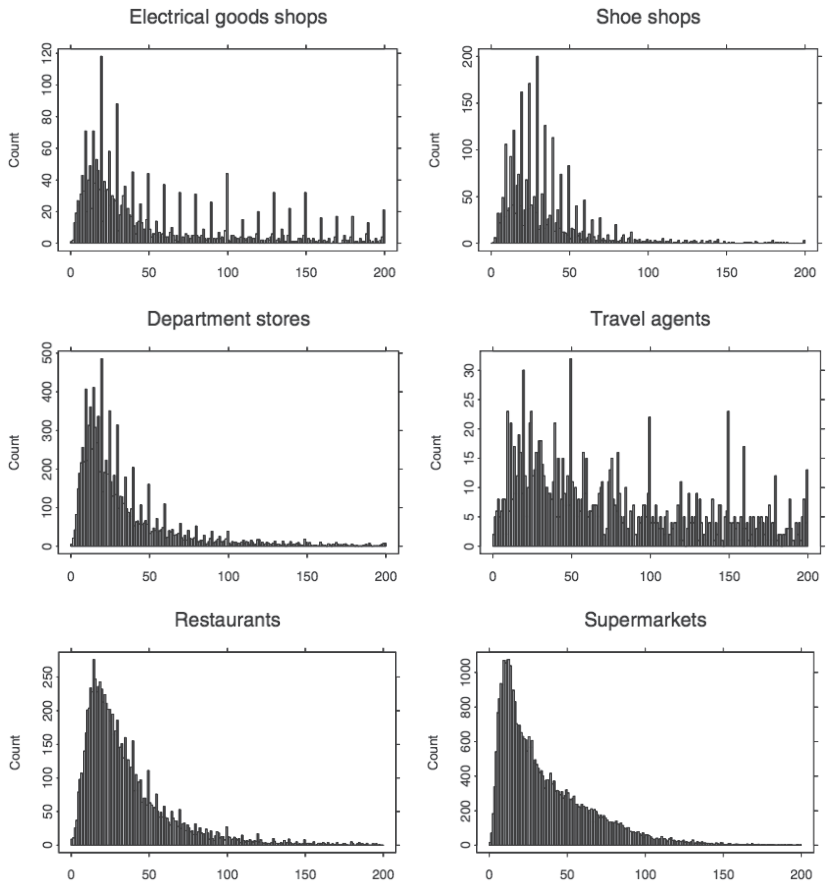
[82] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 239.

[83] Regarding this pattern, it is worth mentioning that "credit card fraudsters often try out a stolen credit card for a low amount to see whether it works, before making a big purchase, resulting in a recent and low monetary value transaction followed by a recent and high monetary value transaction" Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 39.

[84] As Bhattacharyya *et al* explains, "past research suggests that fraudsters try to maximize spending within short periods before frauds get detected and cards are withdrawn". Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Cit.,* p. 603.

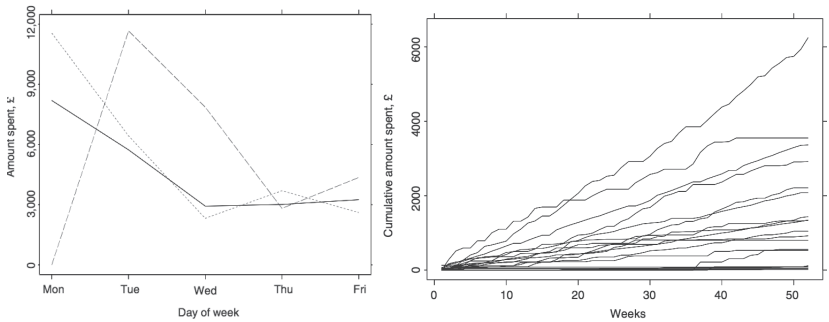
the sudden purchase of numerous goods of high value and that can be easily resold on the black market (namely jewellery or electronic devices), and so on[85].

Figure 4.1. Transaction size distributions for selected trade sectors



[85] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 239. Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 25-26.

Figure 4.2. Transaction patterns for supermarkets



Source: Hand, D. J. and Blunt, G. Prospecting for gems in credit card data. *IMA Journal of Management Mathematics*, 2001, n. 12, pgs. 181, 182 and 185.

Among the types of data used to feed fraud detection algorithms, an important type of aggregated transactional information are the so-called RMF variables, that identify the *recency* (R), that measures the time lapse since the last transaction; the *monetary* aspect (M), notably the minimum, maximum, median and average of historical transactions, as well as the value of the most recent one; and the *frequency* (F) responsible for quantifying the number of transactions made each day, week, month, year and so forth[86]. Besides being useful in detecting credit card misuse alongside other types of fraud[87], the RMF variables are also a powerful technique for identifying and combatting money laundering and terrorist financing, by uncovering patterns typically used by money launderers[88].

As a matter of fact, because a single transaction is very unlikely to appear to be a money laundering event, the application of data mining techniques within the RMF variables help to unveil the three steps associated with money laundering of *placement* (introduction of illegal capital into the banking system), *layering* (undertaking multiple transactions in the legitimate financial system) and *integration* (merging

---

[86] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 39.

[87] As Bart Baesens *at al* points out, the RMF variables "may be operationalized for insurance claim fraud detection by constructing variables such as time since previous claim, number of claims submitted in the previous twelve months, and total monetary amount of claims since subscription of insurance contract". Idem, p. 39.

[88] Idem, p. 39.

the illegal funds with capital from legitimate activities) [89]. One of the ways it does so, for example, is by identifying several transactions made strategically close together but still under the legal threshold by which banks are obligated to report a transaction to authorities – strategy known as smurfing or structuring[90].

Another technique that plays a key role in the fight against money laundering and terrorist financing is *link analysis* through which algorithms can flag transactions as suspicious by confronting them with a database of recorded fraudsters and money launderers, as well as of countries with known connections with terrorist organizations[91]. Finally, when a legitimate company is used to launder money or aid terrorist financing, practices of income statement laundering done by overstating income and expenses can be detect by the analysis of its balance sheet throughout accounting techniques such as Benford's Law test, first two digits test or last two digits tests[92], among others, a theme that shall be addressed in the next section.

V. DATA MINING TECHNIQUES AND FORENSIC ACCOUNTING

The last form of financial fraud against which Big Data and artificial intelligence techniques are applied to prevent and detect are the so-called *financial statement fraud*, a form of misbehaviour by which market participants make false or incomplete statements about the real nature or financial health of a company[93]. Following three main objectives of either covering up the misappropriation or misapplication of funds, misleading investors or regulators about the profitability and the future prospects of the firm or, finally, to facilitate and hide other criminal activities (such as money laundering or tax evasion), this kind

---

[89] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 254-256.

[90] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 239.

[91] Idem, p. 237-239.

[92] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 257-259. Louwers, Timothy J. (2015). The past, present, and future (?) of crime--related forensic accounting methodology. *Cit.*, p. 08.

[93] Reurink, Arjan. Financial fraud: A literature review. *Cit.*, p. 08.

of fraud is committed within the balance sheet of a company by using different forms of accounting manipulation and deception[94]. The association of several harmful consequences caused by these forms of white collar crime in the financial sector and in the real economy with the difficulty of unveiling their occurrence until it is too late, led to the development of the specialized field of forensic accounting[95]. As a merger of both forensic science and accounting, based on the combination of skills and techniques emerging from law, accounting and audit, this field is responsible for the process of assessment, interpretation, summary and presentation of complex financial issues with the main purpose of preventing and detecting fraud[96].

With the increasing volume and complexity of corporate information and the huge amount of structured and unstructured data creating "large samples that will usually be too extensive to review given the auditor or forensic accountant's time constraints"[97], the use of Big Data and artificial intelligence techniques to perform and improve audit and forensic practices is on the rise, in a search for greater operational efficiency[98]. Within the forensic accounting framework, the

---

[94] Young, David. Financial Statement Fraud: Motivation, Methods, and Detection. Baker, H.K., Purda-Heeler, L. and Saadi, S. (Ed.) Corporate Fraud Exposed, Emerald Publishing Limited, Bingley, 2020, p. 325. Regarding the techniques used to commit this type of fraud, Arjan Reurink clarifies that a "review of the literature ... shows that this myriad of techniques can be broken down into five broad categories. The first two of these, *revenue-based schemes* and *expense-based schemes*, aim at artificially boosting a firm's current profitability as reported on the in- come statement. The third and fourth categories, *asset-based schemes* and *liability-based schemes*, involve the fraudulent strengthening of the balance sheet through misrepresentations of asset values and risk exposures, in order to increase a company's financial health and perceived future earnings power. The final category, *other financial statement schemes*, represents a residual one". Reurink, Arjan. Financial fraud: A literature review. *Cit.*, p. 09.

[95] Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Journal of Financial Regulation and Compliance*, 25(1), 2017, p. 73.

[96] Akinbowale, Oluwatoyin Esther; Klingelhöfer, Heinz Eckart; Zerihun, Mulatu Fikadu. An innovative approach in combating economic crime using forensic accounting techniques, *Cit,* p. 1263-1266.

[97] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 620.

[98] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Cit.,* p. 270-271. Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 18.

use of advanced digital tools is increasingly playing a critical role in the steps of data acquisition and management, data analysis and deep investigation and, finally, presentation of findings[99]. During the first phase of preparation of the investigation, for example, not only is it possible to extract files and recover deleted ones, but it is also possible to perform, with enhanced precision and quicker results, a *content examination*, identifying the type of each data file in the system and comparing it with known documents, as well as a *transaction examination*, reviewing the time of its occurrence and the sequence of creation of its data in the system[100].

When it comes to the analysis of large data sets, on the other hand, data mining techniques are a powerful tool for continuous monitoring and periodic analysis, improving the process of transaction testing, proactive fraud detection, detection of abnormalities, unstructured data reviews and pattern recognition[101]. In this context, these data mining techniques can either be *predictive*, used to reduce the risks of fraud in selected business processes, or *descriptive*, employed to detect anomalies in large data sets. While predictive techniques use large historical data sets to predict outcomes of a targeted variable and avoid fraud and manipulation, descriptive ones identify clusters and underlying associations within the data in a search for deviating behaviours that may need further investigation[102].

Even recognizing, following Mark Nigrini, that "[t]he literature lacks a case-based guide for external auditors and forensic accountants to determine which analytics tests might be effective in a proactive fraud detection exercise"[103], it is possible to argue that the use of data mining techniques to analyse the balance sheet of a company is

---

[99] Akinbowale, Oluwatoyin Esther; Klingelhöfer, Heinz Eckart; Zerihun, Mulatu Fikadu. An innovative approach in combating economic crime using forensic accounting techniques. *Cit.,* p. 1259-1260.

[100] Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Cit.,* p. 79.

[101] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 603.

[102] Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Cit.,* p. 78.

[103] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 603.

associated with a wide range of fraud detection tests[104] designed to identify "seven categories of fraudulent number patterns: round numbers, rising numbers, threshold numbers, non-Benford numbers, repeated numbers, outlier numbers, and rounded numbers"[105]. Because the volume of information is usually large, the use of Big Data to perform predefined audit testes combined with a data extraction tool aids the forensic accounting process, allowing the entire data set to be tested and evaluated[106].

Even though the analysis can only provide a list of anomalies and not a set of confirmed fraud cases, the greater analytical capacities provided by these algorithms reveals patterns of interest, reducing notable transactions eligible for a human review to a manageable number of entries which can subsequently be analysed using fraud-audit procedures[107]. Once a fraudulent event is confirmed, the forensic accountant can review its previous hypothesis, adjust its tests and, with a revised plan, perform additional analytical investigations and procedures, in a circular process that may continue several times.

Finally, despite its clear importance, however, this is not the only use of Big Data in the context of forensic accounting practice, as Zabihollah Rezaee and Jim Wang explain[108]:

> First, when forensic accountants investigate fraud, corruption or bribery cases, they take industry-specific norms or regulations into consideration and use keyword phrases to identify potential fraud.

---

[104] For instance, it is possible to mention the Benford's Law Test, Number Duplication Test, Z-Score Test, Relative Size Factor Test, Same-Same-Same Test, Same--Same-Different Test, Trend Analysis, GEL-1 and GEL-2 Tests, Relative Size Factor Test, Even Dollar Amounts, Payments without Purchase Orders Test, Length of Time between Invoice and Payment Dates Test, Payroll Master and Commission Tests. For an overview of each test, see Gee, Sunder. Fraud and fraud detection: a data analytics approach. New Jersey: Ed. Wiley, 2015.

[105] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 603-604. As the author argues, these patterns of fraudulent numbers are associated with three white collar crimes, namely, asset misappropriation, corruption and financial statement fraud.

[106] Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Cit.,* p. 78.

[107] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 67-68.

[108] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Cit.,* p. 270-271.

Second, by using historical activities or transaction data, forensic accountants can use predictive modeling and other advanced analytics to detect suspicious and anomalous transactions, high-risk events, or potential fraudulent behavior or activities. Third, by mining across multiple databases (such as customer or third-party databases), forensic accountants can use entity resolution algorithms to identify hidden relationships, addresses and aliases and investigate conflicts of interest, fake identities or sanctioned individuals and entities. Fourth, forensic accountants use social network analytics to detect hidden relationships, bogus vendors or fake bank accounts when they analyze both structured and unstructured data in the format of visuals and links from social media. Fifth, a large amount of unstructured text data is available from the free text field of journal entries, payment description, expense details, e-mails, social media, documents, presentations and hard drives of individual employees or organizations. Forensic accountants use text mining or text analytics with heuristic rules and statistical techniques to discover the sentiments and conceptual meanings of large amounts of text data, which help to identify potential fraud or non- compliance in the organization. Finally, besides traditional simple spreadsheets or static charts and graphs, forensic accountants use data visualization techniques and interactive dashboards to present evidence in an easy to understand manner.

VI. CONCLUSIONS

The development of new technologies and the digital transformation of society are changing many aspects of life and introducing new ways of social interaction. Alongside the expansion of electronic databases, the development and increased adoption of different IT trends is fostering the evolution and application of Big Data and Artificial Intelligence. The expansion of digital data and the increasing importance of electronic databases for economic development and strategic management are reshaping many aspects of society as well as several different business sectors. The financial services are the leading field in investment in advanced technology for continuous monitoring, transaction analysis, anomaly detection, data review, pattern recognition and other emerging techniques. Simultaneously, the improvement of the data extraction and analysis capacity, enabled by new technological

advances such as data mining and data analysis, is revolutionizing audit and forensic accounting practices and improving the fight against financial crimes, corporate fraud and money laundering.

Data analytics applied to fraud detection uses supervised and unsupervised methods. Supervised methods learn from historical observations from which they extract patterns of fraudulent and non-fraudulent behaviour based on selected samples applied to train algorithms that latter assign a suspicion score to evaluated cases. Unsupervised methods, on the other hand, are trained with a baseline of what represents normal behaviour and focus on detecting anomalies outlining observations that departure from this norm. Along with these methods, social network analysis is also applied by creating a so-called spider construction to analyse network-related information and identify potentially suspicious activities. These three fraud detection techniques are not mutually exclusive and, once each one focuses on a different aspect of fraud, are usually complementary, reinforcing each other's strength and compensating for their vulnerabilities.

The solutions in the market aim at two different types of misconduct associated, firstly, with illegal financial flows in crimes such as credit card fraud, financial identity fraud, money laundering and terrorist financing and, secondly, with different forms of financial fraud, which range from internal corporate crimes, such as money embezzlement and reckless management, to financial statement fraud in the form of market manipulation and investment scams.

Credit card fraud detection algorithms use artificial intelligence and data mining techniques to analyse transactions data as well as its numerical and categorical attributes. It also uses descriptive analytics methods such as *outlier detection techniques* to identify abnormal or anomalous behaviours that might indicate suspicious activities, based on standards of normality obtained by the application of statistical tools, machine learning methods and data mining techniques which examine and identify both individual patterns of previous usage and general patterns of consumption. The so-called RMF variables, that identify the recency (R), monetary aspect (M) and frequency (F) are an important type of aggregated transactional information that are useful in detecting credit card misuse and identifying and combatting money laundering and terrorist financing.

Big Data and artificial intelligence techniques are also applied in order to prevent and detect financial statement fraud by providing

greater operational efficiency for forensic accounting processes, aiding the steps of data acquisition and management, data analysis and deep investigation and, at last, presentation of findings. By being either predictive or descriptive, data mining techniques are a powerful tool for continuous monitoring and periodic analysis, improving the process of transaction testing, proactive fraud detection, detection of abnormalities, unstructured data reviews and pattern recognition. These techniques are used to analyse the balance sheet of a company and perform predefined audit tests designed to identify seven categories of fraudulent number patterns. With its enhanced analytical capacities, data mining techniques reveal patterns of interest, reducing notable transactions eligible for a human review to a manageable number of entries, which can subsequently be analysed using fraud-audit procedures.

REFERENCES

Akinbowale, Oluwatoyin Esther; Klingelhöfer, Heinz Eckart; Zerihun, Mulatu Fikadu. An innovative approach in combating economic crime using forensic accounting techniques, *Journal of Financial Crime*, Vol. 27 n. 4, 2020, p. 1253-1271.

Beasens, Bart. Analytics in a Big Data World. The Essential Guide to Data Science and its Applications. New Jersey: Wiley, 2014.

Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. A Guide to Data Science for Fraud Detection. New Jersey: Wiley, 2015.

Bawack, R.E., Fosso Wamba, S. and Carillo, K.D.A. A framework for understanding artificial intelligence research: insights from practice. *Journal of Enterprise Information Management*, Vol. 34 No. 2, 2021, p. 645-678.

Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50, (3), 2011, p. 602-613.

Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review, Statistical Science, Vol. 17, No. 3 (Aug., 2002), Institute of Mathematical Statistics, pp. 235-249.

European Central Bank. Sixth report on card fraud. August 2020.

Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. "PRECEPT: a framework for ethical digital forensics investigations", *Journal of Intellectual Capital*, 2020, Vol. 21 No. 2, pp. 257-290.

Gee, Sunder. Fraud and fraud detection: a data analytics approach. New Jersey: Ed. Wiley, 2015.

Hand, D. J. and Blunt, G. Prospecting for gems in credit card data. *IMA Journal of Management Mathematics*, 2001, n. 12, 173– 200.

Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Journal of Financial Regulation and Compliance*, 25(1), 2017, p. 73–85.

Lee, Jae-Ung; Soh, Woo-Young. Comparative analysis on integrated digital forensic tools for digital forensic investigation. *IOP Conference Series: Materials Science and Engineering*, 2020, 834(), 012034.

Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology. Accounting Research Journal, 28(1), p. 4–9.

Marr, Bernard. Big Data in Practice. How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. Chichester: Wiley, 2016.

Netherlands Register of Court Experts NRGD, 2016. Standards 008.0 Digital Forensics. Technical Report Netherlands Register of Court Experts. Available at https://www.nrgd.nl/binaries/Standards%20Digital%20Forensics_tcm39-82994.pdf

Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Managerial Auditing Journal*, Vol. 34 No. 5, 2019, p. 602-622.

Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Forensic Science International: Digital Investigation*, Volume 33, 2020, 200908.

Oliveira Júnior, Edson; Zorzo, Avelino F.; Neu, Charles Varlei. Towards a conceptual model for promoting digital forensics experiments. *Forensic Science International: Digital Investigation*, 35, 2020, 301014.

Reurink, Arjan. Financial fraud: A literature review. MPIfG Discussion Paper, No. 16/5, Max Planck Institute for the Study of Societies, Cologne, 2016.

Rezaee, Zabihollah; Wang, Jim. Relevance of big data to forensic accounting practice and education. *Managerial Auditing Journal*, Vol. 34 No. 3, 2019, p. 268-288.

V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, p. 2094-2100.

van Baar, R.B.; van Beek, H.M.A.; van Eijk, E.J. (2014). Digital Forensics as a Service: A game changer. Digital Investigation, 11(), S54–S62.

van Beek, H.M.A.; van den Bos, J.; Boztas, A.; van Eijk, E. J.; Schramp, R.; Ugen, M. Digital forensics as a service: Stepping up the game. *Forensic Science International: Digital Investigation*, Volume 35, 2020, 301021, ISSN 2666-2817.

Young, David. Financial Statement Fraud: Motivation, Methods, and Detection. Baker, H.K., Purda-Heeler, L. and Saadi, S. (Ed.) Corporate Fraud Exposed, Emerald Publishing Limited, Bingley, 2020, p. 321-339.

Young, Michael R. Financial Fraud Prevention and Detection. Governance and Effective Practices. New Jersey: Wiley, 2014.

Wang, Shiguo. A comprehensive survey of data mining-based accounting-fraud detection research. *2010 International Conference on Intelligent Computation Technology and Automation*, ICICTA 2010, 1, 50–53

Wilson, C. (2019), "IBM Tech trends to watch in 2020 . . . and beyond", IBM, available at: https://www.forbes.com/sites/ibm/2019/12/09/ibm-tech-trends-to-watch-in-2020–and-beyond/#5511ae004c1c (accessed on 22nd September 2021).

Wu, Tina; Breitinger, Frank; O'Shaughnessy, Stephen (2020). Digital forensic tools: Recent advances and enhancing the status quo. *Forensic Science International: Digital Investigation*, 34(), 300999.

PART II

RESPONSIBILITY

4

The Last Cocktail - Economic and Financial Crime, Corporate Criminal Responsibility, Compliance and Artificial Intelligence

(https://doi.org/10.47907/livro2021_4c5)

Anabela Miranda Rodrigues[1]

Abstract:
In the current global economic and financial scenario, in which corporations are the main protagonist, the issue of making them and/or their administrators, managers and employees responsible for crimes committed in the business sphere emerges. Compliance has been the "Columbus egg" for regulators and those subject to regulation in recent decades. This statement hides its potentialities and weaknesses, especially when criminal compliance is taken into account, as is the case with this study. Its socializing function is opposed to a security vision of compliance, which recovers the corporation as a «total institution». With AI systems that now combine compliance, it also becomes an "intelligent corporation". Still poorly redone from the trapdoors of vicarious responsibility and ambiguities of the organization defect, finding models of responsibility for corporate's crimes is, for criminal law scholars, again urgent.

Keywords: compliance; intelligent corporation; predictive process; corporate criminal liability.

---

[1] Full Professor of Law. Univ. Coimbra, University of Coimbra Institute for Legal Research, Fac. Law.

I. INTRODUCTION

In the current global economic and financial scenario, in which corporations are the main protagonist, it is not difficult to see how their activity can verge on the criminal, even giving rise to a new phenomenology of it. With their very complex organizational structures and them acting in contexts of increasing risk, the issue of making them and/or their administrators, managers and employees responsible for crimes committed in the business sphere emerges.

Compliance, as a law enforcement strategy and one of the pillars of corporate governance, is assumed as a vector for the assessment of criminal responsibility, and determination of the legal consequences arising from the practice of illicit activities, whose importance varies depending on the model of responsibility adopted by the corporation. In turn, the statement that we are living today in an era of a new business reality made possible not only by the enormous computational development, but above all by Systems of Artificial Intelligence (AI), whose application is enhanced by the enormous development of computing and cognitive communication – the "Internet of Things" (IoT) – will not come as a surprise to anyone. In such a scenario, networked "things" – machines and systems – communicate and interact with each other, showing themselves capable of predicting productive acts and processes in a very effective and efficient way, or preventing or detecting errors harmful to the company. Thus, such an algorithm has the advantage of increasing security in a business context by predicting, preventing and designing harmful acts or values, and monitoring the space and the people who intervene in it. The digital transition also favours the transfer of decisions in the business context to complex computer systems. Partially at least, several options taken throughout the production process are already decided by "things". That is, many of the tasks decided, assigned and previously performed by humans are now assigned to, decided on and performed by machines. However, an erroneous decision by an algorithm, causing injury to legal assets, is in critical conflict with a model of criminal liability built on the performance of a person, human or fictional, or in this case, the legal entity, in any of its models, which we will appreciate in this study.

II. COMPLIANCE - SOCIALIZING FUNCTION AND RISK MANAGEMENT

It is important to remember that the issue of good corporate governance and compliance arises in a most unusual context – that of regulated self-regulation. This involves self-regulation by private entities being subordinated to the purposes and interests of the state. This development means that calls for a need for regulatory intervention are heard ever more loudly, and, in the last resort, these must involve sanctions under the criminal law. Under such a regulatory strategy, the criminal law is like the last guest to arrive at a party, but without whose indispensable presence the festivities cannot start. The purpose of establishing measures of internal organisation of a corporation is not to create a normative programme that favours its activity 'on a knife edge' and allows it to evade criminal liability, but to delimit the perimeter of prohibited conduct, so that practices contrary to the defined rules of conduct can be prevented and suppressed. The possibility of criminal sanctions is a way of encouraging business leaders to establish effective control mechanisms. The motivation to ensure compliance with the control rules is thus the result of corporations overestimating the possibility of non-criminal prosecution and the establishment of procedural agreements or the provision for the exclusion or mitigation of their criminal liability.

In this regard, it should be noted that the compliance strategy, in the light of modern self-regulation, lives with a degree of state intervention different from what it classically was, in this sense, "less co-active and more dialogue". It is a question of focusing intervention, in particular administrative or even criminal, more on the quality and effectiveness of the internal self-regulation system and less, in accordance with the traditional public control model, on the repression of non-compliance with the rule by its addressee. It is a question of avoiding a method of action based on severe sanctions from the outset. In other words: the focus is on preventing corporate misconduct.

In this context, in which compliance is particularly important, the prevention of offence to legal values becomes a duty and a responsibility for corporations and gains a socializing sense – it is the socialization of modern times[2]. Compliance programmes aim to promote an ethical

---

[2] See Rodrigues, Anabela Miranda, *Direito Penal Económico - uma Política na Era Compliance*, Almedina, Coimbra, 2021, 2ª ed., p. 28s.

business culture and legal compliance, and their ultimate objective is to avoid the injury of legal values and the corresponding administrative, civil and ultimately, but above all, criminal liability. This compliance strategy uses a new type of law enforcement in which state action involves introducing a (new) level of law enforcement between the (violation of) the standard and the (application of) sanction or punishment. It is therefore not directed so much to sanction or punish as to "seek the cooperation and participation of infringers, with the aim of correcting the defects that led to the violation of a rule""[3]. In essence, it is a question of making them able to avoid similar behaviours in the future. In the context of business activity, this means that state intervention through compliance fulfils a socializing function.

The effectiveness of compliance thus understood takes into account an aspect that should not be over-ensured. And that lies in the finding that compliance with standards, in the context of the risk in which corporations currently carry out their activity, can involve real difficulties. It is here that the prodigious technological evolution that we are experiencing is felt, by favouring the appearance of algorithms capable of extracting and structuring, from big data, information relevant to business management[4]. One of its most common applications is based on the enormous capacity for business risk assessment, management and control. The most complex deep learning and AI-based technology solutions are of particular importance for their enormous analytical capability and the high capacity of accuracy and anticipation that they are recognized to have. Risk management by the 'machine' covers areas as diverse as the prevention and fight against fraud and the monitoring of the operation of a corporation - acting in the context of product and supplier management or even compliance with legal and regulatory obligations - and of its workers, and several advantages in reducing the enormous costs of regulatory compliance are recognized[5].

---

[3] See Martín, Adán Nieto, "Autorregulación, 'compliance' y justicia restaurativa", *Autorregulación y sanciones, Luis Arroyo Jimenez/Adán Nieto Martin* (Directores), Thomson Reuters, Aranzadi, 2ª Edición, 2015, p.117s (see, also, p.102).

[4] See Rodrigues, Anabela Miranda/ Sousa, Susana Aires, "Algoritmos em contexto empresarial: vantagens e desafios à luz do direito penal", *Julgar*, Nº45, Set-Dez., 2021 (ongoing publication), II, 2., 2.1

[5] See, in a developed way, Butler, Tom / O'Brien, Leona, "Artificial intelligence for regulatory compliance: Are we there yet?", *Journal of Financial Compliance*, Vol. 3, N 1, 2019, p. 44.

In the context of preventing and combating fraud, there are several concrete examples of practical applications that have been developed by financial institutions in order to meet requirements imposed by regulators, for example on money laundering. AI solutions promise continuous monitoring of the company, in turn facilitating the regulator's rapid access to information in the event of non-compliance. *Buttler and O'Brien*[6] refer to a revolution capable of transforming risk and compliance monitoring into a predictive process. The continuous monitoring of the company allows problems to be identified and solved in advance, providing "compliance breaches" and thus preventing the entity regulated (the corporation) from having to answer to the regulator and other judicial authorities. As the organization and analysis of data becomes more targeted and focused through AI, real-time information will enable the self-anticipation of risks and reach the "holy grail" of an intelligent compliance system, as *Aziz* and *Dowling*[7] point out. The prevention and fight against fraud also includes the application of new AI techniques as guarantors of the security and integrity of the financial system, preventing cyberattacks and signalling illegal or criminal situations. The critical software market capable of preventing and detecting fraud is expanding, with more and more companies specializing in the supply of these products. Take the case of *Feedzai*[8], a Portuguese financial technology *start-up*, specializing in fraud detection and cybercrime prevention in the financial and banking sector, using AI and machine learning techniques.

III. COMPLIANCE AND CRIMINAL LIABILITY OF LEGAL PERSONS

1. Compliance relief and intelligent algorithm

Assuming that the sanctioning, inter alia criminal, of much economic and financial behaviour was an overriding necessity, the

---

[6] See Butler /O'Brien, *Journal of Financial Compliance* (note 5), p. 45.

[7] See Aziz, Saqib / Dowling, Michael, "Machine Learning and AI for Risk Management", *Disrupting Finance*, Palgrave Macmillan, 2019, p. 47.

[8] The company has earned media attention for its international valuation of about $1 billion, giving it "unicorn start-up" status. In 2018, *Feedzai* had been considered one of the 50 most promising companies in the field of financial technology by *Forbes*, having received several international distinctions. See https://feedzai.com/about-us/

criminal law faced the first difficulties of accountability in relation to the aggression of legal values in 'collective action contexts[9]. The issue of criminal liability in this criminal field is a significant aspect in the conferral of this responsibility on so-called legal persons. This is what has largely fuelled the doctrinal discussion that has been waged around the possible imputation models of corporate criminal liability. These models can conform to two major systems: the vicarious or heteronomous model, in which the responsibility for the conduct of an administrator, manager or employee is transferred to the collective entity; and the other, which is based on corporate self-responsibility and the possibility of the company being liable for criminal liability for "organisational defect".

It is known that in continental Europe, contrary to the classical theory of criminal law based around the individual agent, an idea of criminal responsibility of collective entities has been established and gradually expanded. If the French criminal law of 1994 and the Belgian law of 1999 are referred to as having enshrined a regime regarded as exceptional and extravagant, it is a fact that the political-criminal solution of the criminal liability of legal persons was deserving of acceptance in criminal codes, even in countries traditionally averse to criminal liability of this nature, such as occurred in 2010 under the Spanish Penal Code. In Portugal, it was in 1984, with legislation regarding infringements against the economy and against public health[10], that the first steps were taken in the establishment of criminal liability of legal persons. Since then, the imputation of criminal responsibility to legal persons has gradually intensified, exponentially increasing the range of crimes that can be committed by them.

It is within an autonomous model of criminal liability that it has been considered that the adoption of compliance programmes can take on importance for corporations. Moreover, today, when considering this relevant fact, the use of intelligent algorithms in the field of self-regulation needs to be taken into account[11]. It is true that a

---

[9] See Sousa, Susana Aires, *Questões Fundamentais de Direito Penal da Empresa*, Almedina, Coimbra, 2019, p. 84s; see also, Rodrigues, Anabela Miranda, note 2, p. 110s.

[10] Law Decree nº28/84, the 20th January.

[11] See Rodrigues, Anabela Miranda/Sousa, Susana Aires, *Julgar* (note 4), III, 1., 1.1.

'smart enterprise' - capable of acting in continuous communication with and impervious to the organisation, to the extent that such defects would be corrected in advance by the algorithm - is still a vision situated in an uncertain future. An algorithmic-based compliance system that automates a company in fulfilling the obligations imposed by regulators, and thus capable of excluding its eventual liability, while an ongoing challenge being tackled by some corporations, is yet to be realized. In legal systems that include models in which the imputation of a criminal act to a legal entity is based on a defect in organization, as happens in Italy or Spain, the "intelligent" compliance software is presented with the promise of being a powerful tool to exclude the legal entity from responsibility, by first of all furnishing the proof that the company organized itself in such a way as to comply with the law. On the corporate side, the advantages of an intelligent compliance system are thus, at first sight, of a dual nature, tangible and normative: the first, concerned with the mitigation or elimination of error and a the consequent increase in security; the second, bringing the business activity closer to a strict regulatory compliance framework capable of excluding the company from any liability.

2. Responsibility of legal persons: again?

There is in general a problematic side to compliance, which translates as the distrust of the justice system in relation to it, considering it an "invention of the business world"[12]. What is said is that the corporations with the greatest bargaining power, large companies, have the increased capacity to convince the criminal investigation bodies – sometimes, with little information and knowledge in these matters – that the system of organisation they have adopted is sufficiently effective to prevent the commission of crimes, and that any crime committed is the result of purely isolated and individual behaviour, of a managing director or employee. From pointing to a scapegoat to avoiding criminal liability is a small step for the corporation. This is especially true for legal regimes that accept corporate self-responsibility or even mixed models of liability. The particularly perverse effect of this strategy,

---

[12] See Rodrigues, Anabela Miranda (note 2), p. 115, with bibliographical references.

known as '"reverse whistleblowing'"[13], is that the company, in order to give consistency to its version of the facts that it is well organized – with cosmetic use of compliance programmes - seeks an individual on whom it can pin the blame. Adoption of such a strategy is additional harmful to the legal system if the collective entity is offered immunity from or mitigation of punishment, or even a non-criminal persecution in exchange for the naming of the individual responsible. In this regard, the paradoxical effect of corporate autonomous criminal liability has been denounced[14] and it is termed an ongoing creation of a 'friend's criminal law' for businesses[15].

As for a model of heteronomous responsibility, failures can be pointed out especially in large companies, where it is more difficult, by virtue of their complexity, to find an individual responsible, and basing the responsibility of the corporation on an action or omission of an individual. To condition the company's responsibility to demonstrate, for example, that any manager of the organisation, in relation to a specific criminal act and a subordinate, has breached his or her supervisory duties, would mean desecrating a model of corporate responsibility that would benefit large companies and harm smaller ones, since in these it is much easier to locate responsibility or the concrete lack of vigilance of a superior, administrator or manager. In any case, this form of imputation of criminal liability to companies - which runs the risk of translating, in judicial practice, into an objective imputation of liability that derives automatically from individual responsibility - promotes a business reaction of concealment of crime and alliance with the offender, which reaches the level of obstruction of justice: the

---

[13] The expression is from Kimberley, D. Krawiec, "Cosmetic Compliance and the Failure of Negotiated Governance F. Hodge O'Neal Corporate and Securities Law Synposium – After the Sarbanes-Oxley Act: The Future of the Mandatory Disclosure System", *Wash U.L.Q.,* 81, 2003, p. 487s.

[14] See Laufer, William S., last, in 2018, "A very special regulatory milestone", *Univ.Pa.J. Bus.Law,* Vol. 20.2., p.391s. See also, Mendes, Paulo Sousa, "*Law Enforcement & Compliance*", *Estudos sobre law enforcement,* Almedina, 2018, p. 26s e Sousa, Susana Aires, *Questões Fundamentais, cit.,* p. 127 e 128

[15] About this, see Rodríguez, Laura Zuñiga, "Responsabilidad penal de las personas jurídicas y derechos humanos. Una valoración desde la reforma de 2015 de la legislación española", *Derecho Penal Económico y Derechos Humanos,* Eduardo Demetrio Crespo, Adán Nieto Martín (Directores), Manuel Maroto Calatayud, Mª Pilar Marco Francia (Coordinadores), Tirant lo blanch, Valencia, 2018, p. 106s.

company is not interested in assisting the investigation, as ultimately its discovery may translate into its conviction. Its fortune is united with that of the person responsible, who turns into its ally.

The introduction of AI into business activity introduces new difficulties to the difficulties already known about from the models of imputation to legal persons, through both individuals and collective individuals.

The issue lies in so-called "intelligent" algorithms, technologically complex, capable of autonomously classifying qualifying options as criminal, but which had not been pre-programmed in this sense even when such decisions were predictable to the programmer (*cognitive robots*)[16]. The novelty is then in the fact that the machine, as a machine that learns", obtains a new result that is, in a sense, its own. As an artificial intelligence system, a "learning machine" must not be confused with a complex data processor, that is, it is not limited to calculating the best option among the thousands of items of data that have been introduced to it, such analysis being inaccessible or very difficult for a human. Rather, the algorithm, powered by data, continually adjusts itself in order to decrease the margin of error and create its own decision. It is this dynamic nature of the machine – its autonomy – that challenges the attribution of responsibility to the people behind the machine, whether physical or legal[17].

It is in this context that the most difficult issues of imputation of corporate criminal liability are identified[18]. In a vicarious model, the question is how to impute the criminally relevant decisions and actions carried out by the machine, under the conditions described, to individuals. In an autonomous model of responsibility, difficulties arise to the precise extent that the "defect" of the algorithm is not

---

[16] On the distinction between cognitive robots and deterministic robots - pre-programmed for the practice of a given criminal activity - clearly, in the context of robots, see *Report of COMEST on Robotic Ethics*, 2017, p. 48, https://unesdoc.unesco.org/ark:/48223/pf0000253952

[17] On the difficulties present here in the area of the imputation of penal responsability, see Sousa, Susana Aires, "'Não fui eu, foi a máquina': teoria do crime, responsabilidade e inteligência artificial", *A Inteligência Artificial no Direito Penal* (coord. Anabela Miranda Rodrigues), Almedina, Coimbra, 2020, p. 65s with bibliographical references.

[18] See Rodrigues, Anabela Miranda/Sousa, Susana Aires, *Julgar,* (note 4), III, 1., 1.1.

known and, as such, preventable and avoidable. The cognitive ability of the machine makes it unpredictable, able to react to the unexpected, and removes its decision from the mastery of predictability of the programmer. It is this space of freedom that is granted to the machine, exploiting its learning abilities, which cannot be determined (or prevented). The "defect" in the algorithm does not exist; it is a future defect and therefore escapes self-organization ... of the algorithm... and thus also the corporation! At least in an abstract sense, if the offence caused by learning of the algorithm leads to an unpredictable outcome, one can hardly blame the corporation for not avoiding a risk it could not know.

At the present time, intelligent business self-organization will not eliminate wrong decisions made by intelligent software, which are proven examples of discriminatory options in hiring or firing workers, price combination situations or phantom financial transactions[19].

In fact, digital transformation of the corporation evidences a patent non-conformity between the technological evolution of corporations and the models legally provided to assess their criminal liability, in turn unveiling a gap already identified by some discourse on the subject. The various proposals for a solution call for an extension or reconfiguration of the assumptions of criminal liability. Faced with the manifest difficulty in making a human, natural, person responsible, the hypotheses oscillate between the modification and updating of the assumptions of corporate responsibility to the most radical ones that propose making the machine responsible.

Referring specifically to this problem, *Mihailis Diamantis* seeks to propose making a corporation responsible, exploring a model that consists of adapting to the business context of "extended mind thesis"[20]. From this perspective, in the process of automating the company, algorithms integrate the way the company thinks and takes decisions and,

---

[19] On the problem involved here and the crimes of market abuse committed by artificial agents, see Rodrigues, Anabela Miranda, "Os crimes de abuso de mercado e a "Escada Impossível" de *Escher* – o Caso do *Spoofing*)", *Julgar*, Nº45, Set.-Dez. 2021 (ongoing publication), *passim*.

[20] Diamantis, Mihailis E., "The Extended Corporate Mind: When Corporations Use AI to Break the Law", 98 N.C. L. Rev. 893 (2020); also, Bryson / Diamantis/ Grant, "Of, for, and by the people: the legal lacuna of synthetic persons", *Art. Intelll Law* (2017), p. 273 e ss.

thus, constitute an extension of its mental state and will, linking it thus with its criminal responsibility.

On the other hand, the supposed insufficiency of the classic legal schemes of attribution of criminal liability have constituted a decisive impulse for the emergence of theoretical proposals that advocate an electronic legal personality, on the civil plane, and a consequent direct criminal liability of the machine as a response to the responsibility / accountability gap. For example, *Gabriel Hallevy* proposes the seemingly simple idea that if the assumptions of criminal liability in an entity are verified, it must be held accountable, be it a physical entity, a collective entity or an artificial entity[21]. In a clear utilitarian understanding of criminal liability, the extension of criminal law to autonomous and intelligent machines would not require, in the author's view, major changes to the assumptions required by this responsibility, it being possible to identify, in the performance of AI, the external (*actus reus*) and mental (*mens rea*) elements required by criminal liability.

IV. CORPORATE CRIME, ACCOUNTABILITY, COMPLIANCE AND AI: THE LAST COCKTAIL

Talking about *compliance* means having in mind the possibility of conceiving two standard models of programmes: one, which may consist of promoting an ethical culture and legality; and another, which is rooted in surveillance or control mechanisms.

Thus[22], according to the first model, the compliance program, whose central element is the ethical code, is oriented towards the promotion of values. It relies, of course, on control measures, which are seen as the normal internal procedures for the operation of a corporation focused on business ethics, namely due diligence, which is

---

[21] Hallevy Gabriel, "The Criminal Liability of Artificial Intelligence Entities – From Science Fiction to Legal Social Control", *Akron Intellectual Property Journal* Vol. 4, Issue 2 (2010), p. 199; *id*, *Liability for crimes involving artificial intelligence systems*, Springer, 2015, p. 61. For a critical assessment of the construction of this author, Sousa, Susana Aires, (note 17), p. 77s. In critical sense, see also Rodrigues, Anabela Miranda, "A justiça preditiva entre a americanização e a europeização", *A Inteligência Artificial no Direito Penal* (coord. Anabela Miranda Rodrigues), Almedina, Coimbra, 2020, p. 52s.

[22] See Rodrigues, Anabela Miranda (note 2), p. 105s.

fundamentally thought of as an instrument for promoting an 'illicit--free' business environment, internalised by all as a reputational added value and in terms of the value of the corporation. In this model, crime reporting is integrated into the corporate culture as a corporate civic act and not with any pejorative connotation[23]. The second is seen as a function of surveillance, and at its heart are control measures. A compliance model from surveillance and control has ingredients - such as using video surveillance circuits, phone records or internet access - that run the risk of converting the company into a kind of *panopticum* and giving the entrepreneur a *big brother* position. In the age of intelligent compliance, perhaps the most appropriate image is that of a "Ubiquitous Digital Architect", of which Soshana Zubof speaks[24]. What's more, criminal compliance becomes a source of misconceptions. The toughening of systems of detection, of reporting, of investigation, the publicity of sanctions (shaming) or the increasing criminalization of many violations of compliance duties criminalize compliance. This new near-criminal law is private. Certainly, the dangers of the privatization of criminal justice are not born out of this new compliance strategy; but it does create new problems.

In this context, it is generally observed that such a model would be incompatible with workers' fundamental rights, such as to a private life or intimacy, the secrecy of communications or the right to data protection. And it's easy to understand how scanning powers this model and powers its costs[25]. The continuous monitoring of workers facilitates

---

[23] The way the reporting channels work is a telling sign of the compliance model deployed. It is essential for an ethical model for channels to be anonymous and specific, enabling administrators and employees and people outside the company to communicate, under conditions of confidentiality, situations that may pose business risks. In this way, it is not necessary to foster an environment of persecution among the staff of the company and of persecution of the staff of the company. And, thus, on the one hand, preventing not only situations of complaints of bad faith, since confidentiality does not prevent the responsibility and sanctioning of the whistleblower, if this is the case; and, on the other hand, seeking to safeguard whistleblowers of good faith communications from disciplinary, professional or criminal repercussions.

[24] See Zuboff, Shoshana, "A Era do Capitalismo de Vigilância. A disputa por um futuro humano na nova fronteira do Poder", Relógio D'Água, 2020, p. 389s.

[25] See Rodrigues, Anabela Miranda/Sousa, Susana Aires, *Julgar* (note 4), III, 2.; see, also, Sousa, Susana Aires, "As diferentes faces dos programas de compliance", Legitimidade e efetividade dos programas de compliance (or. Adán Nieto Martín/Eduardo Saad Diniz), Tirant lo blanch, 2021, p.29s.

the identification of error and, above all, facilitates the pointing out individualized failure of an individual's conduct, identified and indicated by the algorithm. The presumption of liability thus established is added to the double transfer of responsibility from the corporation to individual persons, and among such transfers, from the directors to middle or lower-level management of the corporation (*top-down*). Indeed, the algorithm has the ability to accurately identify the timing of the error, disregarding the context and the "film of the event""[26]. The repercussions at the procedural level, on the presumption of innocence, are evident from this: the "photograph" of the error relieves the company and shifts the burden on to the defence of the worker. The algorithm allows the company to easily overcome the test of the abstract-concrete adequacy of the compliance program by increasing the possibility of excluding its liability at the expense of the presumption of guilt of the worker[27].

V. CONCLUSION

Compliance has been the "Columbus egg" for regulators and those subject to regulation in recent decades. This statement hides its potentialities and weaknesses, especially when criminal compliance is taken into account, as is the case with this study. Its socializing function is opposed to a security vision of compliance, which recovers the corporation as a total institution. With AI systems that now combine compliance, it also becomes an "intelligent corporation". Still poorly redone from the trapdoors of vicarious responsibility and ambiguities of the organization defect, finding models of responsibility for corporate's crimes is, for criminal lawyers, again urgent.

---

[26] See Rodrigues, Anabela Miranda (note 2), p. 112, note 229.

[27] On this issue of particular relevance in autonomous models of criminal liability of companies, Rodrigues, Anabela Miranda, *Direito Penal Económico,* (note 2), p. 112 e s; *id*, "Compliance programmes and corporate criminal compliance", *Polar – Portuguese Law Review,* Vol. 2, January 2018, n.º 1, p. 5s. In the procedural context, it is also important to consider that the algorithm is also a means of obtaining proof, of private creation. In the Portuguese legal order, on the side of the evidential use of this information for the purposes of criminal liability, there will always be the limits insurmountable to its validity, in the light of Article 32(8) of the Constitution and Article 126 of the Code of Criminal Procedure.

REFERENCES

Aziz, Saqib / Dowling, Michael, "Machine Learning and AI for Risk Management", *Disrupting Finance*, Palgrave Macmillan, 2019.

Bryson, j.; Diamantis, M.; Grant, T. D., "Of, for, and by the people: the legal lacuna of synthetic persons", *Art. Intelll Law* 25 (2017), p. 273-291.

Butler, Tom / O'Brien, Leona, "Artificial intelligence for regulatory compliance: Are we there yet?", *Journal of Financial Compliance*, Vol. 3, N 1, 2019.

Diamantis, Mihailis E., "The Extended Corporate Mind: When Corporations Use AI to Break the Law", *98 N.C. L. Rev*. 893 2020.

Hallevy Gabriel, "The Criminal Liability of Artificial Intelligence Entities – From Science Fiction to Legal Social Control", *Akron Intellectual Property Journal* Vol. 4, Issue 2 (2010).

--- *Liability for crimes involving artificial intelligence systems*, Springer, 2015.

Kimberley, D. Krawiec, "Cosmetic Compliance and the Failure of Negotiated Governance F. Hodge O'Neal Corporate and Securities Law Synposium – After the Sarbanes-Oxley Act: The Future of the Mandatory Disclosure System", *Wash U.L.Q.,* 81, 2003.

Laufer, William S., "A very special regulatory milestone", *Univ.Pa.J. Bus.Law,* Vol. 20.2., 2018.

Martín, Adán Nieto, "Autorregulación, 'compliance' y justicia restaurativa", *Autorregulación y sanciones, Luis Arroyo Jimenez/Adán Nieto Martin* (Directores), Thomson Reuters, Aranzadi, 2ª Edición, 2015.

Mendes, Paulo Sousa, "*Law Enforcement & Compliance*", *Estudos sobre law enforcement,* Almedina, Coimbra, 2018.

Rodrigues, Anabela Miranda, "Compliance programmes and corporate criminal compliance", *Polar – Portuguese Law Review,* Vol. 2, January 2018, n.º 1.

---- "A justiça preditiva entre a americanização e a europeização", *A Inteligência Artificial no Direito Penal* (coord. Anabela Miranda Rodrigues), Almedina, Coimbra, 2020.

---- *Direito Penal Económico: Uma Política Criminal na Era Compliance*, 2ª Ed, Almedina, Coimbra, 2021.

---- "Os crimes de abuso de mercado e a "Escada Impossível" de *Escher* – o Caso do *Spoofing*)", *Julgar*, Nº45, Set.-Dez. 2021 (ongoing publication).

Rodrigues, Anabela Miranda/ SOUSA, Susana Aires, "Algoritmos em contexto empresarial: vantagens e desafios à luz do direito penal", *Julgar*, Nº45, Set-Dez., 2021 (ongoing publication).

Rodríguez, Laura Zuñiga, "Responsabilidad penal de las personas jurídicas y derechos humanos. Una valoración desde la reforma de 2015 de la legislación española", *Derecho Penal Económico y Derechos Humanos,* Eduardo Demetrio Crespo, Adán Nieto Martín (Directores), Manuel Maroto Calatayud, Mª Pilar Marco Francia (Coordinadores), Tirant lo blanch, Valencia, 2018.

Shoshana Zuboff, "A Era do Capitalismo de Vigilância. A disputa por um futuro humano na nova fronteira do Poder", Relógio D'Água, 2020.

Sousa, Susana Aires, *Questões Fundamentais de Direito Penal da Empresa*, Almedina, Coimbra, 2019.

----, "'Não fui eu, foi a máquina': teoria do crime, responsabilidade e inteligência artificial", *A Inteligência Artificial no Direito Penal* (coord. Anabela Miranda Rodrigues), Almedina, Coimbra, 2020.

----, "As diferentes faces dos programas de compliance", Legitimidade e efetividade dos programas de compliance (or. Adán Nieto Martín/Eduardo Saad Diniz), Tirant lo blanch, 2021.

5

Algorithmic Harms as Corporate Misconduct

(https://doi.org/10.47907/livro2021_4c6)

Mihailis E. Diamantis[1]

Abstract
Algorithms offer many social benefits, but when they discriminate in lending, manipulate stock markets, or violate expectations of privacy, they can injure us on a massive scale. The problem is that algorithms fit poorly into existing conceptions of liability. Liability requires injurious acts, but what does it mean for an algorithm to act? This Chapter offers a solution. Corporations currently design and run the algorithms that have the most significant social impacts. Corporate law stipulates that corporations act through their employees because corporations have control over and benefit from employee conduct. This Chapter argues that the same control and benefit rationales could extend to corporate algorithms. If the law were to recognize that algorithmic conduct qualifies as corporate action, the existing framework of corporate liability would engage when corporate algorithms cause harm.

Keywords: Al injury; corporate liability; respondeat superior

---

[1] Associate Professor, The University of Iowa College of Law. I am grateful to the George Washington Law Review for its permission to write this chapter, which is an abridged version of my article, *Algorithms Acting Badly: A Solution from Corporate Law*, 81 Geo. Wash. L. Rev. (forthcoming 2021).

[A] robot may not injure a human being or, through inaction, allow a human being to come to harm.
— Isaac Asimov, The First Law of Robotics[2]

I. THE LEGAL CHALLENGE OF ALGORITHMIC INJURY

The first law of robotics is already dead. Robots and the algorithms that run them injure people every day. Some of these injuries are tragically palpable. For example, in 2015, an assembly robot at a car plant in Ionia, Michigan bypassed safety protocols, entered an unauthorized area, and crushed employee Wanda Holbrook's head. In 2018, a self-driving car struck and killed pedestrian Elaine Herzberg as she was walking across the street in Tempe, Arizona. Some algorithmic injuries are less visceral, but are just as disruptive because they impact thousands of people. Algorithms that extend loans or hire employees often discriminate against minority applicants.[3] Stock-trading algorithms capable of executing thousands of trades a second can artificially distort stock prices for higher profit.[4] Price-setting algorithms from competing retailers can collude to raise costs for customers.[5]

When robots and algorithms injure people (whether physically, financially, or otherwise), recovery and justice can prove elusive. Many forms of criminal and civil liability require that (or are much easier to prove if) someone directly harms another. In cases of algorithmic harm, the algorithm stands between the victim and any legally cognizable defendant. Wanda Holbrook's husband struggled in his case to find a suitable defendant. Prosecutors decided they could not press charges against Uber for killing Elaine Herzberg. Victims of algorithmic

---

[2] Isaac Asimov, *Runaround*, *in* I, Robot 25, 37 (Bantam Books 2004) (1950).

[3] *See* Robin Nunn, *Discrimination and Algorithms in Financial Services: Unintended Consequences of AI*, Cyberspace Law., Apr. 2018, at 4, 4 (discussing "AI's so called 'white guy problem'").

[4] Enrique Martínez-Miranda, Peter McBurney & Matthew J. Howard, *Learning Unfair Trading: A Market Manipulation Analysis from the Reinforcement Learning Perspective*, Ass'n for Advancement A.I. (2015), https://arxiv.org/pdf/1511.00740.pdf [https://perma.cc/J226-GPTD]; Tom C.W. Lin, *The New Market Manipulation*, 66 Emory L.J. 1253, 1284–85 (2017).

[5] Michal S. Gal, *Algorithms as Illegal Agreements*, 34 Berkeley Tech. L. J. 67 (2019).

discrimination flounder about for a theory of liability.[6] Algorithmic stock manipulation is hard to prosecute unless there is a guilty human pulling the strings.[7] And antitrust law has yet to see its first case alleging purely algorithmic collusion.[8]

There are compelling reasons to use algorithms. Although some take lives, they have the capacity to save many more. Although some discriminate in lending or hiring, they have the potential to make these processes more objective. Although some manipulate markets, effective algorithmic trading can also make markets more efficient. We have only scratched the surface of the cost savings and big-data insights that robots and algorithms will come to offer. These social benefits, however, are no guarantee that algorithms will not harm us along the way. Most experts are skeptical that advanced algorithms are worth the risk. The fact is, "[a]s robotics and artificial intelligence systems increasingly integrate into our society, they will do bad things."[9]

The key to making algorithms work for us, rather than against us, is to use the law to address the threats they pose. Accountability is the law's most direct and effective tool for turning behavior in socially constructive directions. And yet there is currently no general framework for algorithmic accountability. In reporting on Elaine Herzberg's death, a journalist hit on the central challenge: "Who killed Elaine Herzberg? Not the driver of the car that ran her over — because there was no driver. And therein lies a problem."[10] When people kill each

---

[6] *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Calif. L. Rev. 671, 711–12, 726 (2016).

[7] *See generally* Lin, *supra* note 3, at 1300–01.

[8] The closest have been cases that involve algorithms purposely developed by competing retailers to collude on pricing. *See, e.g.*, Andrew C. Finch, Acting Assistant Att'y Gen., Dep't of Just., Antitrust Div., Remarks at the 44th Annual Conference on International Antitrust Law and Policy (Sept. 14, 2017), https://www.justice.gov/opa/speech/file/996756/download [https://perma.cc/2RKN-8ZKV].

[9] Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 U. Chi. L. Rev. 1311, 1311 (2019).

[10] Angie Schmitt, *Uber Got Off the Hook for Killing a Pedestrian with its Self-Driving Car*, Streetsblog (Mar. 8, 2019), https://usa.streetsblog.org/2019/03/08/uber-got-off-the-hook-for-killing-a-pedestrian-with-its-self-driving-car/ [https://perma.cc/6BDN-6X7Y]. There was a human "monitor" in the car. Jack Stilgoe, *Who Killed Elaine Herzberg?*, Medium: OneZero (Dec. 12, 2019), https://onezero.medium.com/who-killed-elaine-herzberg-ea01fb14fc5e [https://perma.cc/YBA8-RXTT]. The monitor seems to have been looking down (perhaps at her phone) at the

other or manipulate stock, the law knows how to respond. When algorithms do the same, there can be a wide gap in legal accountability.

To close the algorithmic accountability gap, the law needs to say what liability looks like when algorithms are behind the wheel. Most liability, whether criminal or civil, requires an injurious act. Acts are the sorts of things that only people can do, but algorithms are not people.

Scholars in law,[11] computer science,[12] and business ethics[13] who have broached the question of algorithmic liability often assume that the solution is to recognize algorithms as people. However, granting algorithms the status of legal persons is deeply unappealing for several reasons. First, it would require a seismic reworking of current law; algorithms are presently not legal people and they cannot be civil or criminal defendants. Even were that to change, there is no way to sanction algorithms: they lack bodies to jail and pocketbooks to pay.[14]

More worryingly for the sci-fi readers out there, it would be foolhardy to assume that the slick slope of algorithmic personhood stops

---

time of the crash. *Id.* Attention fatigue for human monitors in self-driving cars is a natural and predictable event. *See* Jack Stewart, *Self-Driving Cars Won't Just Watch the World—They'll Watch You*, Wired (Feb. 13, 2017, 7:30 AM), https://www.wired.com/2017/02/self-driving-cars-wont-just-watch-world-theyll-watch/ [https://perma.cc/UW2D-FEKT]. In Uber's eyes, this only made it easier to distance the company, morally and legally, from the tragedy: "[W]e refused to take responsibility. They blamed it on the homeless lady, the Latina with a criminal record driving the car.... But *our car* hit a person. No one inside [Uber] said, 'We did something wrong and we should change our behavior.'" Julie Bort, *Uber Insiders Describe Infighting and Questionable Decisions Before Its Self-Driving Car Killed a Pedestrian*, Bus. Insider (Nov. 19, 2018, 5:17 PM), https://www.businessinsider.com/sources-describe-questionable-decisions-and-dysfunction-inside-ubers-self-driving-unit-before-one-of-its-cars--killed-a-pedestrian-2018-10 [https://perma.cc/H8UY-WMP5].

[11] *See* Gabriel Hallevy, Liability for Crimes Involving Artificial Intelligence Systems 7 (2015).

[12] *See* Luciano Floridi & J.W. Sanders, *On the Morality of Artificial Agents*, 14 Minds & Machs. 349, 350–51 (2004).

[13] *See* Nicholas Diakopoulos & Sorelle Friedler, *How to Hold Algorithms Accountable*, MIT Tech. Rev. (Nov. 17, 2016), https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/ [https://perma.cc/78NG-BEFS].

[14] *See* Ryan Abbott & Alex Sarch, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, 53 U.C. Davis L. Rev. 323, 364–68, 383 (2019); Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. Rev. 1231, 1244–48 (1992).

with liability. Rights usually accompany responsibilities in law,[15] and the prospect of pitting algorithm rights against human rights is full of chillingly unanticipatable consequences.[16] We have seen this dynamic play out before with other artificial persons. Could the early engineers of legal personhood for corporations a century ago have predicted the conflict in the United States between corporations and individuals for religious freedom and political speech?

There is a silver lining to the cautionary tale of corporate personhood — whatever its faults, it is here to stay, and, as argued below, it may offer a scaffold for constructing a legal response to algorithmic injuries. There was no legally responsible natural person driving the car that killed Elaine Herzberg. There was no legally responsible algorithm driving the car either, because algorithms, not being people, cannot be responsible. The basic thesis advanced here is that there was a third possibility, an overlooked person in control of the car: Uber.

Corporations develop, run, and maintain the world's most impactful algorithms. Just as corporations act through their employees, I argue below that they may also act through their algorithms. Holding corporations liable for the things they do through their employees induces corporations to ensure that their employees behave in socially beneficial ways. Recognizing that corporations act through their algorithms would similarly encourage corporations to exercise responsible control over algorithmic injuries. By converting the question of injurious algorithmic action into a question of injurious corporate action, the approach advanced here crucially avoids the practical and philosophical challenges that accompany any effort to personify algorithms. Algorithms become an extension of the corporate person, not persons in their own right.

In pursuit of realistic prospects for success, it grounds itself in existing corporate law and the principles behind it. Part II details the current law of corporate liability in the United States, emphasizing how the law conceives of injurious corporate action by looking for

---

[15] *See* W. Robert Thomas, *How and Why Corporations Became (and Remain) Persons Under the Criminal Law*, 45 Fla. St. U. L. Rev. 479, 504–14 (2018).

[16] *See* Joanna J. Bryson, Mihailis E. Diamantis & Thomas D. Grant, *Of, for, and by the People: The Legal Lacuna of Synthetic Persons*, 25 A.I. & L. 273, 275 (2017) (criticizing the possibility of extending rights to algorithms in part because of the implications it would have for humans' rights).

an injurious employee action to attribute to the corporation. Part III shows how law, as presently applied, cannot close the algorithmic accountability gap because algorithmic injury has no obvious place in it.

Part IV argues that an approach to algorithmic accountability may be hiding in plain sight. The principles behind the current law of corporate liability — which emphasize relationships of control and benefit — extend beyond the employment context. This Chapter offers a "beneficial-control account" according to which a corporation could be liable for an algorithmic injury if it claims the substantial productive benefits of the algorithm and exercises sufficient control over it.

As Part V shows, recognizing that corporations act through algorithms just as they act through employees would go a long way to address algorithmic injury. This would establish a responsible party against whom victims could seek satisfaction. And that, in turn, would incentivize corporations to take care to discipline their algorithms by designing, releasing, monitoring, and updating them responsibly. Though there would be some challenges with implementation, Part V shows they are surmountable. Part VI concludes and notes some limitations of using corporate law to solve the algorithmic accountability gap.

II. THE LAW OF CORPORATE LIABILITY

The law of liability was built with human defendants in mind. Liability typically requires some kind of injurious act — e.g., driving over someone. We intuitively understand what it means for human defendants to act.

Corporations are different. "A corporation is an artificial being, invisible, intangible, and existing only in contemplation of law."[17] In order for corporations to fulfill their economic and social role, there must be some sense in which they are capable of doing things. "[A] corporation must of course be able to act... [or] else the whole theory of incorporation would make no sense whatsoever."[18] They need to

---

[17] Trs. of Dartmouth Coll. v. Woodward, 17 U.S. (4 Wheat.) 518, 636 (1819).

[18] Gerhard O.W. Mueller, *Mens Rea and the Corporation: A Study of the Model Penal Code Position on Corporate Criminal Liability*, 19 U. Pitt. L. Rev. 21, 38 (1957).

purchase property, set up factories, make goods, and intend to bind themselves to agreement in order to participate meaningfully in the marketplace. There is no intuitive sense of what corporations are or what it means for a corporation to act. So the law had to define it.

Lawmakers took two crucial shortcuts in defining corporate action. Because the law was creating an entirely new entity, it could have developed a parallel legal system from scratch, defining afresh what legal concepts mean as applied to corporations. Instead, they took the first shortcut, slotting corporations into existing law just as if they were other "people."[19] As the Supreme Court has observed, "the corporate personality is a fiction, although a fiction intended to be acted upon as though it were a fact."[20] Accordingly, any statute that defines civil or criminal liability for people simultaneously creates a cause of action applicable to individuals and to corporations.

Simply declaring that corporations are people who can violate a law says nothing about how to tell when a violation has occurred. Since corporations do not have physical bodies, the law had to define how corporations act.[21] This challenge prompted lawmakers to take a second shortcut. Rather than turn to some sophisticated, policy-driven approach tailored to the corporate context, lawmakers simply pilfered a doctrine from an ancient law that applied to Roman slaveholders: respondeat superior.[22] That doctrine attributed the misdeeds of slaves to their owners. Transposed to the corporate context, respondeat superior now largely means that corporations "do" whatever their employees do.[23]

---

[19] *See* 1 U.S.C. § 1 ("In determining the meaning of any Act of Congress, unless the context indicates otherwise... the words 'person' and 'whoever' include corporations... as well as individuals.").

[20] Int'l Shoe Co. v. Washington, 326 U.S. 310, 316 (1945).

[21] Mihailis E. Diamantis, *The Body Corporate*, 83 L. & Contemp. Probs. 133 (2020).

[22] *See, e.g.*, Phila., Wilmington, & Balt. R.R. Co. v. Quigley, 62 U.S. (21 How.) 202, 209–10 (1859). Some trace the doctrine as far back as Roman times. *See* Oliver Wendell Holmes, Jr., *Agency*, 4 Harv. L. Rev. 345, 350 (1891).

[23] Though there is some additional nuance. The employees have to be working "within the scope of their employment" for their thoughts and acts to be attributable to the corporation; however, the employee satisfies this condition even if she is disobeying orders. United States v. Hilton Hotels Corp., 467 F.2d 1000, 1004 (9th Cir. 1972). Employees must also have some intent to benefit the corporation to attribute their acts and thoughts, though they satisfy this condition even if their intent is

III. IS ANY CHANGE NEEDED?

In this Part, I consider whether present law, more creatively applied, could close the algorithmic accountability gap. Perhaps respondeat superior could work if judges were to focus in a more sophisticated way on the conduct of employees who design corporate algorithms. Or perhaps employees were the wrong place to look in the first place; if corporations make algorithms, maybe principles drawn from products liability could close the gap. In the two Sections that follow, I argue that, as they presently stand, neither body of law is sufficient.

A. Respondeat Superior

Designing, training, and running algorithms presently requires human involvement. Humans write the code, compile the data sets, and train the algorithms.[24] If algorithmic behavior ultimately traces back to human acts, then perhaps respondeat superior's identification of corporate acts with human acts limits courts less than expected. Maybe courts just need to understand more about how algorithms are made and how to locate the cause of algorithmic injury in deficiencies of responsible corporate programmers.

Creative use of respondeat superior is not nearly enough to close the algorithmic accountability gap because there are, and increasingly will be, many algorithmic injuries that have no direct connection to human employees. Today, algorithms for the most part originate with human engineers; however, humans are increasingly absent from the process. There once was a time when humans needed to write every line of code, but now algorithms themselves write most of the code for sophisticated programs.[25] Humans are still usually involved — they generally supervise the process — yet even now there are techniques

---

subsidiary, United States v. Automated Med. Lab'ys, Inc., 770 F.2d 399, 407 (4th Cir. 1985), hypothetical, United States v. Sun-Diamond Growers of Cal., 138 F.3d 961, 970 (D.C. Cir. 1998), *aff'd*, 526 U.S. 398 (1999), and ineffective, *see Old Monastery Co.*, 147 F.2d at 908.

[24] *See* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. Davis L. Rev. 653, 668 (2017).

[25] *See* Catherine Tremble, Note, *Wild Westworld: Section 230 of the CDA and Social Networks' Use of Machine-Learning Algorithms*, 86 Fordham L. Rev. 825, 837 (2017).

for unsupervised algorithmic learning.[26] As humans have less and less of a hand in the process of software development, the applicability of respondeat superior to algorithmic conduct becomes increasingly tenuous.

Even today, where software engineers have a heavy hand in supervised algorithmic learning, respondeat superior is often inadequate for closing the algorithmic accountability gap. To see why, it is important to understand the type of corporate algorithms that are most concerning. The most powerful algorithms today are not the mechanistic if--A-output-B programs of yesteryear and freshman computer science courses. Those algorithms required technicians to write every line of code, to anticipate every possible input, and to specify every possible output. The algorithms that hold the most promise for boosting corporate productivity largely design themselves using a technique called "machine learning."[27] After specifying a machine learning algorithm's goal, programmers train it with a set of test cases, telling the algorithm in each instance whether or not it attained its goal.[28] With each test case, the algorithm updates its own code and eventually learns how to perform the task on its own. The result is a program that, at least in many respects, can accomplish a goal faster, more accurately, and cheaper than any human. It is also an algorithm that no human could have designed from the ground up; the resulting code is often inscrutable, so complicated that no one reading it afterwards can understand how it works.

Because machine learning code is often effectively a black box, algorithms can behave in ways that are unintended, unexpected, and unpredictable by any human intelligence. This is part of the power of machine learning. Employees who do precisely as their employers command are less valuable than employees who can interpret commands with a dose of common sense and flexibly apply them to changing circumstances. The same is true of algorithms. Machine

---

[26] Jason Brownlee, *Supervised and Unsupervised Machine Learning Algorithms*, Mach. Learning Mastery (Mar. 16, 2016), https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/ [https://perma.cc/YAA9-CM49].

[27] *See* Lemley & Casey, *supra* note 88, at 1335 ("[T]he unpredictability inherent in machine learning is also one of its greatest strengths.").

[28] *See* Lehr & Ohm, *supra* note23, at 668.

learning is so powerful precisely because it moves beyond the basic code its programmers are capable of writing. This is helpful because the algorithms will solve problems in ways human programmers could not anticipate. But if algorithms behave in unforeseeable ways, they will also sometimes do things that employers, and the law, prefer they would not.

Creative use of respondeat superior to triangulate between corporations, their employee programmers, and their algorithms is not a general solution to the algorithmic accountability gap. Machine learning raises the possibility that algorithms will misbehave without any intervening human misconduct.[29] Because machine learning algorithms effectively program themselves, they can draw unanticipated conclusions from test data and interact with the real world in unforeseeable ways. Technologists widely recognize that smart algorithms can misbehave even if every human involved is fully innocent.[30] Without human misconduct, respondeat superior's vision of corporate misconduct cannot apply.

B. Product Liability

There are some mechanisms for imposing corporate liability that — unlike respondeat superior — do not require employee misconduct. One of the best known of these mechanisms is civil products liability. Regardless of what any employee did or thought, when a product's manufacturing or design defect leads to injury, the corporation that made the product is liable.[31] Requiring tort claimants to prove that some employee at some point in the design or manufacturing process did something negligent would present a prohibitive evidentiary barrier. Accordingly, products liability is strict — it requires no misconduct on the part of the corporation or its employees. Could products liability close the algorithmic liability gap? Holding corporations strictly liable for their algorithmic injuries could be an elegant way to sidestep the whole problem of locating and attributing an injurious act.

---

[29] Pedro Domingos, The Master Algorithm 5 (2015).

[30] Barocas & Selbst, *supra* note 5, at 729.

[31] *See* Restatement (Second) of Torts § 402A (Am. L. Inst. 1965) ("[Strict products liability applies even though] the seller has exercised all possible care in the preparation and sale of his product....").

Products liability has several limitations that disqualify it from being an effective way to address algorithmic injury. Perhaps most fundamentally, many of the algorithms that hurt people are not "products." A product is "[s]omething that is distributed commercially for use or consumption."[32] Although the software on a self-driving car sold to consumers probably qualifies, the software that hedge funds use to execute automatic trades or that banks use to make lending decisions certainly do not. Such programs may be developed in-house for corporate use rather than distribution.

Even if algorithms qualify as "products," a further limitation of products liability enters the fray — products liability only clearly applies when there is "physical harm... to the ultimate user or consumer, or to his property."[33] "Casual bystanders, and others who may come in contact with the product, as in the case of employees of the retailer... or a pedestrian hit by an automobile, have been denied recovery."[34] Most people harmed by algorithms, like Elaine Herzberg and Wanda Holbrook and people who face algorithmic discrimination in lending, are not consumers of the algorithms that hurt them.

IV. ALGORITHMIC CORPORATE CONDUCT

Algorithms themselves are not people under the law and so are not themselves subject to suit. Although most algorithms are developed, owned, and operated by corporations, those corporations are also often immune from suit because algorithmic injuries do not fit into respondeat superior's employee-focused vision of corporate misbehavior. Trying to restrain corporate use of algorithms is not a viable path forward because the future of economic development and corporate progress lies in algorithms. At the same time, the course of that development and progress should not be charted over the bodies and livelihoods of the victims of algorithmic injury. We need a way to reliably insert some accountability into the landscape, to recompense victims, and to discipline those who profit from algorithms.

---

[32] *Product*, Black's Law Dictionary (11th ed. 2019).
[33] Restatement (Second) of Torts § 402A(1).
[34] *Id.* § 402A cmt. o.

The law already has a template for responding to the algorithmic accountability gap. More than a century ago, it confronted a structurally similar issue that arose in the wake of large-scale employment. Just like algorithms, employees sometimes injure people in ways that their corporate employers cannot predict. Suing the employees as individuals was an ineffective response because employees usually lack adequate personal resources to make victims whole.[35] Additionally, identifying responsible individuals within corporations is often an insurmountable difficulty.[36] As a policy matter, focusing exclusively on employees as potential defendants also overlooks the common criminogenic role of corporate-level systems and ethos.[37]

The law's solution was to deem that employee acts count as acts of their corporate employer. This gave victims and prosecutors another potential defendant from whom to seek justice. It also gave corporations some skin in the game when their defective systems enabled or encouraged employee misconduct.[38] This incentivizes corporations to train, monitor, and discipline their employees better.

A similar development could work for algorithmic injuries. To define a new type of corporate conduct — algorithmic corporate conduct — the law must say when an algorithm does something on the corporation's behalf. A path forward emerges if one abstracts from the particular application of respondeat superior in the employment context to appreciate the deeper corporate law principles behind the doctrine. As explained in the Sections that follow, these are principles about corporate control (of employees) and corporate benefit (from employees). Respondeat superior's basic requirements provide guidelines for courts to ensure that, for an employee to qualify as acting for a corporation, she should be under the corporation's control and

---

[35] Richard Frankel, *Regulating Privatized Government Through § 1983*, 76 U. Chi. L. Rev. 1449, 1455 (2009).

[36] Memorandum from Eric Holder, Deputy Att'y Gen., to All Component Heads and U.S. Att'ys 4 (June 16, 1999), http://www.justice.gov/sites/default/files/criminal-fraud/legacy/2010/04/11/charging-corps.PDF [https://perma.cc/4ELA-QNMN].

[37] Cindy R. Alexander & Mark A. Cohen, *The Causes of Corporate Crime: An Economic Perspective*, *in* Prosecutors in the Boardroom 11, 17 (Anthony S. Barkow & Rachel E. Barkow eds., 2011); Fiona Haines, Corporate Regulation 25 (1997).

[38] *See* Mihailis E. Diamantis, *Successor Identity*, 36 Yale J. on Reg. 1, 18, 24–25 (2019).

benefiting the corporation.[39] The next two Sections explore principles of control and benefit to say what they might mean as applied to algorithms rather than human employees. The third Section draws both principles together to propose a unified test for when a corporation acts through an algorithm.

A. A Control-Based Account

Deterrence and prevention are some of the most important goals of civil and criminal corporate liability. Corporations are in the best position to address the harms they cause because they have the most information about those harms and have the greatest power to shape the underlying causal mechanisms.[40] By threatening corporations with penalties when those harms result, the law hopes it can induce corporations to exercise their influence over those mechanisms in socially productive ways.[41]

With respect to employees as potential sources of corporate harm, deterrence is an important justifying premise for respondeat superior.[42] From its beginning, courts explained the rationale behind respondeat superior by reference to the "control" that employers exercise over their employees.[43] By holding employers liable for the behavior of their employees, respondeat superior presses employers to use that control to steer employees away from misconduct.[44] Employers have many tools at their disposal for shaping employee behavior, such as commands, incentives, monitoring, training, and discipline. Because employers interact with their employees on a daily basis and establish the context in which productive or destructive business behavior takes place, they are in a unique position to determine how employees behave.[45]

---

[39] Restatement (Third) Of Agency § 7.07(2) (Am. L. Inst. 2006).

[40] Holder Memo, *supra* note 35, at 2.

[41] Larry D. Thompson, *The Blameless Corporation*, 47 Am. Crim. L. Rev. 1251, 1255 (2010).

[42] Harvey L. Pitt & Karl A. Groskaufmanis, *Minimizing Corporate Civil and Criminal Liability: A Second Look at Corporate Codes of Conduct*, 78 Geo. L.J. 1559, 1573 (1990).

[43] *See* Holmes, *supra* note 21, at 347.

[44] Albert W. Alschuler, *Two Ways to Think About the Punishment of Corporations*, 46 Am. Crim. L. Rev. 1359, 1380 (2009).

[45] *See* Fleming James, Jr., *Vicarious Liability*, 28 Tul. L. Rev. 161, 168 (1954).

If corporate liability is about getting corporations to prevent harms that are under their control, there is no reason to limit its reach to employee misconduct. There are other sources of harm that corporations are in a privileged position to manage. A *control-based account* of corporate action would recognize as corporate acts any effects over which a corporation exercises substantial control. As applied to algorithmic injuries, the control-based account would apply whenever an algorithm causes an injury that a corporation had the substantial power to prevent. Just as corporations can fire employees, they can pull the plug on computer programs. Although nothing can guarantee that a machine learning algorithm will always follow the law, there are steps corporations can take to reduce the probability that the algorithm will cause harm.[46] These steps include diversifying the body of engineers writing algorithms,[47] more careful initial programming,[48] more mindful selection of training data sets,[49] more extensive pre-rollout testing,[50] regular post-rollout quality audits,[51] routine run-time compliance layers,[52] effective

---

[46] *See generally* William D. Smart, Cindy M. Grimm & Woodrow Hartzog, An Education Theory of Fault for Autonomous Systems (2017) (describing ways to reduce educational failures in algorithms), http://people.oregonstate.edu/~smartw/library/papers/2017/werobot2017.pdf [https://perma.cc/J2LD-ZCZ6].

[47] *See* Kate Crawford, Opinion, *Artificial Intelligence's White Guy Problem*, N.Y. Times (June 25, 2016), https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html [https://perma.cc/5ZTR-GR74].

[48] *See* Mark A. Geistfeld, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, 105 Calif. L. Rev. 1611, 1634–36 (2017).

[49] Oscar H. Gandy, Jr., *Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems*, 12 Ethics & Info. Tech. 29, 30 (2010).

[50] Dave Cliff & Linda Northrop, *The Global Financial Markets: An Ultra-Large-Scale Systems Perspective*, *in* Large-Scale Complex IT Systems 29, 29 (Radu Calinescu & David Garlan eds., 2012).

[51] James Guszcza, Iyad Rahwan, Will Bible, Manuel Cebrian & Vic Katyal, *Why We Need to Audit Algorithms*, Harvard Bus. Rev. (Nov. 28, 2018), https://hbr.org/2018/11/why-we-need-to-audit-algorithms [https://perma.cc/WA3D-M3FV].

[52] *See* Felipe Meneguzzi & Michael Luck, *Norm-Based Behaviour Modification in BDI Agents*, 8 Int'l Conf. on Autonomous Agents & Multiagent Sys. 177, 177–78 (2009), https://dl.acm.org/doi/pdf/10.5555/1558013.1558037 [https://perma.cc/2PYV-NDS2]; Louise Dennis, Michael Fisher, Marija Slavkovik & Matt Webster, *Formal Verification of Ethical Choices in Autonomous Systems*, 77 Robotics & Autonomous Sys. 1, 2–3 (2016).

monitoring,[53] and continuous software updates to address problems as they arise.[54] Each of these precautions entail costs that, all things considered, corporations would rather avoid. Through the threat of sanction, the law can make taking precaution cheaper than risking violation.

To make the control-based account workable in practice, the law would need to specify several indicia of control to guide factfinders at trial. These indicia should be powers that tell in favor of finding that the corporation had the requisite control. Measuring corporate control over algorithms requires a multifaceted approach because the relationship between corporations and algorithms is not always straightforward. One corporation may design the algorithm, a second may own it, a third may use it, a fourth may own the hardware that runs the algorithm, and a fifth may monitor and update it.[55] Algorithmic injuries could trace to any of those five contributions or to an interaction between them. Trying to measure corporate control over algorithms by using a simple proxy — e.g., which corporation designed the algorithm, which owns it, or which uses it — risks missing the mark where the proxies overlap and intersect in complex ways. The law would do better to inquire directly about corporate power over algorithms.

The relevant powers are those that confer the ability to prevent algorithmic injury. These include the power to design the algorithm in the first place, the power to pull the plug on the algorithm, the power to modify it, and the power to override the algorithm's decisions. A corporation need not have these powers directly in order to count as possessing them. For example, a corporation may have indirect power if it has the legal or economic influence to induce another corporation to act. None of these powers standing alone is determinative of corporate control over algorithms, but the more powers a corporation has, the more control it has. It may even happen that more than one

---

[53] Thomas C. King, Nikita Aggarwal, Mariarosaria Taddeo & Luciano Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, 26 Sci. & Eng'g Ethics 89, 110-11 (2019).

[54] *See* Nat'l Highway Traffic Safety Admin., Federal Automated Vehicles Policy 16 (2016), https://www.hsdl.org/?view&did=795644 [https://perma.cc/S9R-V-KH8L].

[55] *See* Andrew Tutt, *An FDA for Algorithms*, 69 Admin. L. Rev. 83, 106 (2017); Marta Infantino & Weiwei Wang, *Algorithmic Torts: A Prospective Comparative Overview*, 29 Transnat'l L. & Contemp. Probs. 309, 353 (2019).

corporation has control, in which case injurious algorithmic conduct may be attributable to multiple corporate defendants.

Standing alone, the control-based account is ultimately unappealing because it risks expanding the scope of corporate liability for algorithmic injuries too far. Consider, for example, a corporation that operates a social media platform. The corporation may exhibit all of the indicia of control over the platform: it may have designed the platform and have the powers to pull it down, regularly modify it, and override anything the platform does. Even if the corporation exercises its control responsibly, users may end up manipulating features of the platform in ways that injure third parties, perhaps by sending offensive messages, violating intellectual property, or engaging in identity theft. In these sorts of cases, it would be inappropriate to automatically hold the corporation responsible, despite its control over the algorithms that run the platform. Pursuing prevention against corporations too vigilantly risks dampening innovation.[56] Especially when it comes to the fast-developing digital space, domestic corporations must be able to innovate if they are to remain competitive with foreign peers and to deliver the social value that algorithms promise.[57]

B. A Benefits-Based Account

In the law of corporate liability, fairness is an enduring concern, but because corporations are not typical moral agents, it can be difficult to comprehend what "fairness" means as applied to them. Oftentimes, shareholder interests are substituted for corporate interests, and fairness toward corporations translates to fairness toward shareholders.[58] From a fairness perspective, corporate liability is an odd development. The strong presumption is usually that holding one person to account for injuries caused by another person, violates basic fairness norms.[59]

---

[56] Rebecca Crootof, *The Internet of Torts: Expanding Civil Liability Standards to Address Corporate Remote Interference*, 69 Duke L.J. 583, 663 (2019.

[57] *See* Gustavo Manso, *Creating Incentives for Innovation*, 60 Cal. Mgmt. Rev. 18, 18 (2017).

[58] *See* Deborah A. DeMott, *Beyond Metaphor: An Analysis of Fiduciary Obligation*, 1988 Duke L.J. 879, 917.

[59] Scales v. United States, 367 U.S. 203, 224–25 (1961); Dan B. Dobbs, Paul T. Hayden & Ellen M. Bublick, The Law of Torts § 425 (2d ed. 2011); Shawn Bayern, *Three Problems (and Two Solutions) in the Law of Partnership Formation*, 49 U. Mich. J.L. Reform 605, 622–23 (2016).

Corporate liability is vicarious at two different levels. At one level, corporate liability transmits burdens vicariously to individuals from corporations. Though the law may formally punish or award damages against corporations, it can do this only by way of forcing corporations' shareholders to pay.[60] The most powerful response to this fairness-based concern is that the burdens of corporate misconduct for shareholders come paired with the benefits of corporate success.[61] Because shareholders participate in the upside of corporate gains, it is fair for them to share in the losses when things go awry and third parties get hurt.[62]

At a second level, corporate liability also transmits fault vicariously to corporations from individuals. Because corporations can only misbehave through employees, respondeat superior holds corporations to account for the misconduct of employees.[63] At this level too, the most powerful fairness rationale has to do with pairing burdens with benefits: because corporate employers enjoy the benefits of employees' productive activity, they should share in its burdens too.[64] "Just as liability for damage can be equitably balanced against the defendant's fault, so it can be equitably balanced against his benefit."[65] This is part of the rationale behind respondeat superior's requirement that an employee intend to benefit her employer — it limits the doctrine to those cases where employer benefits are to be expected.

Pairing the burdens of productive activity with its benefits mitigates the fairness concerns that arise by allocating burdens or benefits separately. Once again, the logic behind respondeat superior applies

---

[60] *See* John C. Coffee, Jr., *"No Soul to Damn: No Body to Kick": An Unscandalized Inquiry into the Problem of Corporate Punishment*, 79 Mich. L. Rev. 386, 401 (1981); Barnali Choudhury & Martin Petrin, Corporate Duties to the Public 194 (2019).

[61] Restatement (Second) of Agency § 219 cmt. a (Am. L. Inst. 1958) ("[I] t would be unjust to permit an employer to gain from the intelligent cooperation of others without being responsible for the mistakes, the errors of judgment and the frailties of those working under his direction and for his benefit.").

[62] *See* Sara Sun Beale, *A Response to the Critics of Corporate Criminal Liability*, 46 Am. Crim. L. Rev. 1481, 1484–85 (2009).

[63] *See* Larry May, *Vicarious Agency and Corporate Responsibility*, 43 Phil. Stud. 69, 71 (1983).

[64] T. Baty, Vicarious Liability 32 (1916).

[65] Glanville Williams, *Vicarious Liability and the Master's Indemnity*, 20 Mod. L. Rev. 220, 230 (1957).

beyond the employment context. Looking beyond employees to other sources of corporate benefit motivates a *benefits-based account* according an algorithmic injury is attributable to any corporation that claims the substantial benefits of the algorithm.

Like the control-based account, the benefits-based account is an unappealing solution to the algorithmic accountability gap. Although its underlying logic is fairness, it threatens to extend to cases where fairness and sound policy would call for a different result. Consider a simple example. Estimates of how much Google's search engine makes off each individual user range from $10.09 up to $359.00. By contrast, some economists estimate that the average user of internet search services like Google values them at $17,500.00. So individual users claim the vast majority of the productive benefit of search algorithms like Google. Yet, as a matter of fairness or preventive policy, it would make very little sense to hold the otherwise innocent third parties that use web search services liable (and to let Alphabet off) when Google search injures someone, e.g., by facilitating illegal access to copyrighted material or making illegal use of protected personal information.

C. The Beneficial-Control Account

The control-based and benefits-based accounts each speak to different values in the law of corporate liability: prevention and fairness, respectively. They also offer very different criteria for determining when algorithmic injury should qualify as a corporate act. Trying to choose between the control-based and benefits-based accounts presumes a false dichotomy between prevention and fairness. There is no reason the law should have to choose — it should instead demand both. A *beneficial-control account* would accomplish this by treating attributing algorithmic injuries to corporations only when both the control--based and benefits-based criteria are met. This would ensure that each imposition of corporate liability for algorithmic misconduct satisfies both preventive and fairness constraints. Indeed, respondeat superior is a version of a beneficial-control account limited just to employees. The doctrine requires that employees acted within the scope of their employment (a rough proxy for corporate control) *and* intended to benefit their corporate employer (a rough proxy for corporate benefit).

Just as employees routinely satisfy the control-based and benefits-based criteria, so will algorithms. One obvious reason is that corporate

control generally begets corporate benefit. Corporations are rational, profit-seeking enterprises. So they will turn any resource they control to their benefit. An unproductive employee will be retrained. An unprofitable corporate algorithm will be modified. Those resources and mechanisms that corporations cannot turn to their benefit are generally not within their control or will not be for long. Corporations fire wayward employees. They discontinue incorrigible algorithms.

Even though many algorithms will routinely satisfy the beneficial-control criteria, there are many instances in which they will not. Importantly, the benefits-based criteria constrain the most concerning overbreadth of the control-based criteria, and vice versa. Recall the example of the control-based account's overbreadth — a social media platform fully controlled by a corporation but put to illegal and injurious ends by a user. Assuming the corporation is not also profiting from the illegal use, then this case would fail the benefits-based criteria. Similarly, the example above of the benefits-based account's overbreadth involved a user that benefits from a third-party search engine. If the search engine ends up causing injuries, it would make no sense to hold the user liable. Fortunately, the beneficial-control account can accommodate this result because the user would not satisfy the control-based criterion.

As test cases, we might inquire how the beneficial-control account would address the cases of Wanda Holbrook and Elaine Herzberg, with which the Chapter began. Recall that a robot escaped and killed Wanda Holbrook in the manufacturing plant where she worked and a self-driving car killed Herzberg. For both, justice proved elusive because of the algorithmic accountability gap: the law had no straightforward way to recognize the algorithmic conduct as the sort of corporate action to which liability could attach.

There is no question in both cases that Ventra Ionia — the manufacturer that Holbrook worked for — and Uber — which owned the car that ran over Herzberg — claimed substantial benefit from the productive activity of the algorithms at issue. As to control, Uber seemed to satisfy all the indicia for its self-driving cars, which it designed, monitored, and modified, and which it could terminate or override. For Ventra Ionia, the control analysis is more nuanced and would depend on additional facts, which are not publicly available. It does not seem that Ventra Ionia designed the robot that killed Holbrook. It is also

unclear whether Ventra Ionia had the power to implement any modifications or could have shut down the robot or overridden its behavior when it attacked. If Ventra Ionia lacked these indicia of control, there would be no case under the beneficial-control account for saying that Ventra Ionia killed Holbrook through its robot. Instead, perhaps the corporation that designed the robot or had the power to monitor, update, and shut down the robot could be another potential defendant.

The beneficial-control account seems to check all the boxes for an appealing solution to the algorithmic accountability gap. To begin, it identifies a potential class of defendants from whom victims of algorithmic misconduct may seek redress. In so doing, the account also embraces both of the major values that corporate liability should serve: prevention and fairness. By imposing criteria responsive to both control-based and benefits-based concerns, it cabins the overbreadth that either set of criteria would have on its own.

V. EVALUATING THE BENEFICIAL-CONTROL ACCOUNT

The corporate law solution to the algorithmic accountability gap proposed here mirrors existing law. It does for algorithmic misconduct what respondeat superior does for employee misconduct—it opens space for holding corporations accountable. By imposing scope of employment and intent-to-benefit constraints on when employee action is attributable to corporations, respondeat superior effectively asks first whether a corporation had control over and could expect to benefit from employee activity. The beneficial-control account extends this inquiry to the algorithmic context by treating algorithmic activity as corporate action only when the corporation has control over and claims the benefits of the algorithm. This gives the beneficial-control account several attractive advantages over the current state of the law and competing proposals. Still, some challenges linger. I address them below.

A. Advantages

By slotting itself into the existing law of corporate liability, the beneficial-control account offers a comprehensive solution to the algorithmic accountability gap. Most other proposals discuss only narrow categories of algorithmic injury, like self-driving car accidents,

discrimination in hiring, and stock fraud. The law already has well-developed mechanisms for holding corporations liable for all manner of civil and criminal violations. By translating algorithmic injury into a species of corporate misconduct, the present proposal leverages that existing law to cover every recognizable form of algorithmic injury.

The beneficial-control account has several advantages that are familiar from discussions of respondeat superior. By attributing algorithmic injuries to corporations when the corporations are in control of the algorithms, the beneficial-control account makes good on its preventive ambitions. A corporation that exercises control over an algorithm is in the best position to design it carefully to reduce the risk of injury, monitor its performance for injuries it may be causing, modify its code to prevent the injury from recurring, and, if necessary, pull the plug. By requiring that corporations claim the substantial benefits of an algorithm before attributing the algorithmic activity to the corporation, the law would stand by its commitments to fairness and justice. Pairing benefits with liabilities ensures that the costs of algorithmic injury fall where they can best be borne, both financially and morally.

Indeed, the familiarity of the beneficial-control account is one of its chief advantages. The few other comprehensive proposals for closing the algorithmic accountability gap would require dramatic reimagining of existing law (e.g., developing a mechanism for "punishing robots")[66] or wholesale creation of new law (e.g., developing a new fiction of algorithmic personhood).[67] These proposals are long on grandiose vision, but they are short on realistic prospects. Respondeat superior is judge-made law, and its expansion into the law of corporate liability has largely been a judge-led process. If, as argued here, the same principles that motivated respondeat superior in the first place could justify its extension to algorithms, judges just might spring for it.

The beneficial-control account departs from the structure of respondeat superior in one important respect. Respondeat superior generally applies both to corporate acts and corporate mental states. The beneficial-control account limits itself to acts. This is important for two reasons. First, it opens the possibility of adopting a more

---

[66] Christina Mulligan, *Revenge Against Robots*, 69 S.C. L. Rev. 579, 592 (2018).

[67] *See* Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 Harv. J.L. & Tech. 353, 399 (2016).

defensible account of corporate fault. The beneficial-control account only says when algorithmic injuries are attributable to a corporation. That is generally not enough to hold a corporation liable. Ordinarily, before imposing liability, the law also requires that the defendant was somehow at fault, as evidenced by a culpable mental state accompanying the injury. By near universal agreement, respondeat superior is a very poor measure of corporate fault. Better proposals are available, some of which are tailored to the algorithmic context. The second reason it is important that the beneficial-control account only attributes actions and not fault is that it avoids the perils of strict liability. By also requiring that genuine corporate fault, however measured, accompany algorithmic injury, the beneficial-control account strikes a balance between potential corporate defendants and potential plaintiffs. It caters to the public's interests in innovation and recompense, without giving decisive and paralyzing preference to either. Lawmakers already struck this equilibrium by requiring fault in the first place. The beneficial--control account seeks to preserve that equilibrium.

B. Challenges

The beneficial-control account faces two main challenges. The first regards implementation. As discussed above, the inquiry into whether a corporation exercised beneficial control over an algorithm is fact intensive. Uncovering and introducing evidence that pertains to the various indicia of control over and monetization of an algorithm will require a significant commitment of resources from litigants and courts.[68] This is complicated by the fact that multiple corporations may exercise different types of control over or claim different benefits from the same algorithm. Furthermore, applying the control and benefit tests requires drawing lines in grey areas to determine when the control exercised and the benefits claimed are "substantial" enough for liability. This sort of vagueness injects a fair measure of unpredictability into the process that brings its own costs to litigants, both present and prospective.[69]

---

[68] Infantino & Wang, *supra* note 54, at 354.

[69] *See* Andrew Morrison Stumpff, *The Law Is a Fractal: The Attempt to Anticipate Everything*, 44 Loy. U. Chi. L.J. 649, 676 (2013); Richard A. Posner, *Savigny, Holmes, and the Law and Economics of Possession*, 86 Va. L. Rev. 535, 565 (2000).

Any attempt to trivialize these litigation and uncertainty costs would be disingenuous; however, they must be juxtaposed with the costs of alternatives. The challenge is to navigate the perennial tension between easier to implement, bright-line rules and harder to implement, vague standards.[70] Rules are predictable but inflexible. They can, at best, only roughly correlate to more complex underlying economic or justice values that the law seeks to promote.[71] This means that rules will inevitably dictate counterproductive results where they fail to track the subtler contours of value. Standards, by contrast, are less predictable but more flexible, which allows the law to hew more closely to its goals.[72] The decision between applying a rule or a standard turns on how the rule's costs of error compare to the standard's uncertainty and administrative costs.[73] Sometimes, as in strict products liability, rules are preferable for weighing corporate liability.[74] In other cases, lawmakers have decided that standards make more sense — e.g., by requiring "proximate causation" for tort claims against corporations, by requiring "reckless disregard" in workplace safety suits, and by evaluating corporate books for "reasonable assurances" against foreign bribery.

There are various possible rule-like alternatives to the beneficial--control test, but they entail unacceptably high costs that the beneficial-control test avoids. One possible approach is to maintain the status quo, which effectively dictates that algorithmic injury in itself can never qualify as corporate action. In this Chapter, I argued extensively against the present law, which effectively immunizes corporations against liability for algorithmic injuries unless there is some culpable human employee in the loop. This limits corporations' incentives to ensure their algorithms are safe and encourages them to move hastily

---

[70] *See generally* Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 Duke L.J. 557, 562–67 (1992).

[71] *Id.* ("[Rules produce] an imperfect fit... resulting in some outcomes that are erroneous from the standpoint of the substantive principle....").

[72] Kathleen M. Sullivan, *Foreword: The Justices of Rules and Standards*, 106 Harv. L. Rev. 22, 66 (1992).

[73] Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 Harv. L. Rev. 1685, 1689 (1976).

[74] *See* David G. Owen, *Rethinking the Policies of Strict Products Liability*, 33 Vand. L. Rev. 681, 684–85 (1980).

toward automation as a risk management strategy.[75] When corporations can externalize the costs of an activity that otherwise benefits them, we should expect them to do so. This leaves victims without recourse, effectively subsidizing corporate profits with victims' injured bodies, pocketbooks, and dignity.

Rule-like alternatives that would modify the status quo would entail different, but equally disqualifying costs. I have already mentioned the possibility that the law could hold corporations strictly liable for the injuries their algorithms cause. This approach, however, risks unduly depressing algorithmic innovation, which could permanently handicap domestic economic development vis-à-vis foreign competitors. A strict liability approach is also an incomplete solution. In a world where algorithmic development, ownership, licensing, use, and modification are all carried out by different corporate actors, a strict liability approach must still determine on *whose behalf* an algorithm acts. In a sense, then, a strict liability account just passes the buck on a question that the beneficial-control account answers directly.

Somewhere between all (the strict liability approach) and nothing (the status quo) are various possible rule-like variations of the beneficial-control test. It is possible that substantial control in the test could be replaced with one or two prespecified indicia of control, and substantial benefit could be replaced with a bright-line dollar threshold. The concern here is that any effort at line drawing will be an immediate invitation to corporate gamesmanship that would defeat the whole purpose of modifying the status quo. Powers over and monetization of an algorithm can be parceled out in an indefinite number of ways; motivated corporate actors are sure to find ways to retain effective control and benefit while sidestepping any bright-line rule. Additionally, the space of algorithmic innovation is evolving so fast that it is doubtful any rigid legal test would remain relevant for long. A multifactored standard like the beneficial-control test has the flexibility to evolve alongside technological developments.

---

[75] *See* Mihailis E. Diamantis, *The Problem of Algorithmic Corporate Misconduct*, N.Y.U. Program on Corp. Compliance & Enf't: Compliance & Enf't (Sept. 16, 2019), https://wp.nyu.edu/compliance_enforcement/2019/09/16/the-problem-of--algorithmic-corporate-misconduct/ [https://perma.cc/AJW5-52V2].

CONCLUSION

In the coming years, the algorithmic accountability gap will grow to a chasm unless the law takes proactive measures to close it. The stories of Elaine Herzberg and Wanda Holbrook will not remain one-off parables of law's inability to deliver justice. Whether we are prepared to recognize it or not, algorithms have injured us all by distorting stock markets, engaging in anticompetitive collusion, misusing personal information, and discriminating against us. The law must find some sweeping accountability mechanism for algorithmic injury if it is to have any chance of protecting us in the coming age of automation.

This Chapter has focused on one obstacle the law must overcome to close the algorithmic accountability gap: figuring out how to fit algorithms into the existing liability regime. Liability requires injurious action, but algorithms are not agents or people under the law, so the concept of action is inapplicable. The solution proposed here adapts fixtures of corporate law to the algorithmic context. Although algorithms are not legal people capable of acting, corporations are. Today's most impactful algorithms are closely tied to the corporations who develop and use them for their own ends. If the law were to recognize that corporations can act through their algorithms, it would not matter that algorithms are incapable, in the eyes of the law, of acting alone. Injuries caused by corporate algorithms would become injuries caused by corporate action. The victims of those injuries could then seek justice from the corporations who control and profit from the algorithms.

The proposed "beneficial-control account" treats algorithmic injury as a species of corporate action when the corporation has control over and seeks to benefit from the underlying algorithm. This gives victims a potential corporate defendant from whom to seek justice. When a corporation controls an algorithm, the potential for liability will encourage it to exercise greater care in designing, monitoring, and modifying the algorithm going forward. This will result in fewer algorithmic injuries. When a corporation seeks to benefit from the algorithm, holding the corporation accountable is fair even though doing so will otherwise burden innocent corporate stakeholders.

Before closing, I should note one important limitation of the beneficial-control test. Although it can go a long way to closing the algorithmic accountability gap today and for the foreseeable future, there

are possible long-term developments that would necessitate further legal change. By drawing on corporate law and its extensive liability framework, the beneficial-control account presumes, as is largely the case today, that a corporation is behind every significant algorithm. Technologists and science fiction authors envision a future world where this may not be true, where algorithms are self-forming, self-executing, and operate under the control and for the benefit of no one. The freestanding, autonomous algorithm raises what some have called the "hard" problem of algorithmic accountability because there is no one, corporate or natural, to hold to account in the algorithm's stead.[76] In such a future, the beneficial-control test would be of little help. The law needs a solution to the algorithmic accountability gap now, and the beneficial-control account offers an approach suited to circumstances as they exist today. If the algorithmic accountability gap reopens in the future, we will know what that future looks like when it arrives and will be in a better place to develop a solution suited to those times. At that point, some of the proposals that I set aside in this Chapter, like the possibility of recognizing algorithms as legal persons, may no longer seem so far-fetched.

REFERENCES

Abbott, Ryan / Sarch, Alex, Punishing Artificial Intelligence: Legal Fiction or Science Fiction, 53 U.C. DAVIS L. REV. 323, 364–68, 383 (2019).

Alexander, R. / Cohen, Mark A., The Causes of Corporate Crime: An Economic Perspective, in PROSECUTORS IN THE BOARDROOM 11, 17 (Anthony S. Barkow & Rachel E. Barkow eds., 2011).

Alschuler, Albert W., Two Ways to Think About the Punishment of Corporations, 46 AM. CRIM. L. REV. 1359, 1380 (2009).

Barocas, Solon / Selbst, Andrew, D. Big Data's Disparate Impact, 104 CALIF. L. REV. 671, 711–12, 726 (2016).

Bayern, Shawn, Three Problems (and Two Solutions) in the Law of Partnership Formation, 49 U. MICH. J.L. REFORM 605, 622–23 (2016).

---

[76] Abbott & Sarch, *supra* note 13.

Beale, Sara Sun, A Response to the Critics of Corporate Criminal Liability, 46 AM. CRIM. L. REV. 1481, 1484–85 (2009).

Brownlee, Jason, Supervised and Unsupervised Machine Learning Algorithms, MACH. LEARNING MASTERY (Mar. 16, 2016).

Bryson, Joanna J. / Diamantis, Mihailis E. / Grant, Thomas D., Of, for, and by the People: The Legal Lacuna of Synthetic Persons, 25 A.I. & L. 273, 275 (2017).

Choudhury, Barnali / Petrin, Martin, Corporate Duties to the Public 194 (2019).

Cliff Dave / Northrop, Linda, The Global Financial Markets: An Ultra-Large-Scale Systems Perspective, in LARGE-SCALE COMPLEX IT SYSTEMS 29, 29 (Radu Calinescu & David Garlan eds., 2012).

Coffee, Jr., John C., "No Soul to Damn: No Body to Kick": An Unscandalized Inquiry into the Problem of Corporate Punishment, 79 MICH. L. REV. 386, 401 (1981).

Crootof, Rebecca, The Internet of Torts: Expanding Civil Liability Standards to Address Corporate Remote Interference, 69 DUKE L.J. 583, 663 (2019).

DeMott, Deborah, A. Beyond Metaphor: An Analysis of Fiduciary Obligation, 1988 DUKE L.J. 879, 917.

Dennis, Louise / Fisher, Michael / Slavkovik, Marija / Webster, Matt Formal Verification of Ethical Choices in Autonomous Systems, 77 ROBOTICS & AUTONOMOUS SYS. 1, 2–3 (2016).

Diakopoulos, Nicholas / Friedler, Sorelle, How to Hold Algorithms Accountable, MIT TECH. REV. (Nov. 17, 2016), https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/ [https://perma.cc/78NG-BEFS].

Diamantis, Mihailis E., Successor Identity, 36 YALE J. ON REG. 1, 18, 24–25 (2019).

____, The Body Corporate, 83 L. & CONTEMP. PROBS. 133 (2020).

____, The Problem of Algorithmic Corporate Misconduct, N.Y.U. PROGRAM ON CORP. COMPLIANCE & ENF'T: COMPLIANCE & ENF'T (Sept. 16, 2019), https://wp.nyu.edu/compliance_enforcement/2019/09/16/the-problem-of-algorithmic--corporate-misconduct/ [https://perma.cc/AJW5-52V2].

Dobbs, Dan B. / Hayden, Paul T. / Bublick, Ellen M., THE LAW OF TORTS § 425 (2d ed. 2011).

Domingos, Pedro, THE MASTER ALGORITHM 5 (2015).

Floridi, Luciano / Sanders, J.W., On the Morality of Artificial Agents, 14 MINDS & MACHS. 349, 350–51 (2004).

Frankel, Richard, Regulating Privatized Government Through § 1983, 76 U. CHI. L. REV. 1449, 1455 (2009).

Gal, Michal S., Algorithms as Illegal Agreements, 34 BERKELEY TECH. L. J. 67 (2019).

Gandy, Jr., Oscar H., Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems, 12 ETHICS & INFO. TECH. 29, 30 (2010).

Geistfeld, Mark A., A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation, 105 CALIF. L. REV. 1611, 1634–36 (2017).

Guszcza, James / Rahwan, Iyad / Bible, Will / Cebrian, Manuel / Katyal, Vic, Why We Need to Audit Algorithms, HARVARD BUS. REV. (Nov. 28, 2018), https://hbr.org/2018/11/why-we-need-to--audit-algorithms [https://perma.cc/WA3D-M3FV].

Haines, Fiona, CORPORATE REGULATION 25 (1997).

Hallevy, Gabriel, LIABILITY FOR CRIMES INVOLVING ARTIFICIAL INTELLIGENCE SYSTEMS 7 (2015).

Holmes, Jr., Oliver Wendell, Agency, 4 HARV. L. REV. 345, 350 (1891).

Infantino, Marta / Wang, Weiwei, Algorithmic Torts: A Prospective Comparative Overview, 29 TRANSNAT'L L. & CONTEMP. PROBS. 309, 353 (2019).

James, Jr., Fleming, Vicarious Liability, 28 TUL. L. REV. 161, 168 (1954).

Kaplow, Louis, Rules Versus Standards: An Economic Analysis, 42 DUKE L.J. 557, 562–67 (1992).

Kennedy, Duncan, Form and Substance in Private Law Adjudication, 89 HARV. L. REV. 1685, 1689 (1976).

King,Thomas C. / Aggarwal, Nikita / Taddeo, Mariarosaria / Floridi, Luciano, Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions, 26 SCI. & ENG'G ETHICS 89, 110-11 (2019).

Lehr, David / Ohm, Paul, Playing with the Data: What Legal Scholars Should Learn About Machine Learning, 51 U.C. DAVIS L. REV. 653, 668 (2017).

Lemley, Mark A. / Casey, Bryan, Remedies for Robots, 86 U. CHI. L. REV. 1311, 1311 (2019).

Manso, Creating Incentives for Innovation, 60 CAL. MGMT. REV. 18, 18 (2017).

Martínez-Miranda, Enrique / McBurney, Peter / Howard, Matthew J., Learning Unfair Trading: A Market Manipulation Analysis from the Reinforcement Learning Perspective, ASS'N FOR ADVANCEMENT A.I. (2015), https://arxiv.org/pdf/1511.00740.pdf [https://perma.cc/J226-GPTD]; Tom C.W. Lin, The New Market Manipulation, 66 EMORY L.J. 1253, 1284–85 (2017).

May, Larry, Vicarious Agency and Corporate Responsibility, 43 PHIL. STUD. 69, 71 (1983).

Meneguzzi, Felipe / Luck, Michael, Norm-Based Behaviour Modification in BDI Agents, 8 INT'L CONF. ON AUTONOMOUS AGENTS & MULTIAGENT SYS. 177, 177–78 (2009), https://dl.acm.org/doi/pdf/10.5555/1558013.1558037 [https://perma.cc/2PYV-NDS2].

Mueller, Gerhard O.W., Mens Rea and the Corporation: A Study of the Model Penal Code Position on Corporate Criminal Liability, 19 U. PITT. L. REV. 21, 38 (1957).

Mulligan, Christina, Revenge Against Robots, 69 S.C. L. REV. 579, 592 (2018).

Nunn, Robin, Discrimination and Algorithms in Financial Services: Unintended Consequences of AI, CYBERSPACE LAW., Apr. 2018.

Owen, David G., Rethinking the Policies of Strict Products Liability, 33 VAND. L. REV. 681, 684–85 (1980).

Pitt, Harvey L. / Groskaufmanis, Karl A., Minimizing Corporate Civil and Criminal Liability: A Second Look at Corporate Codes of Conduct, 78 GEO. L.J. 1559, 1573 (1990).

Posner, Richard A. / Savigny, Holmes, and the Law and Economics of Possession, 86 VA. L. REV. 535, 565 (2000).

Scherer, Matthew U., Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, 29 HARV. J.L. & TECH. 353, 399 (2016).

Smart, William D., Grimm, Cindy M. / Hartzog, Woodrow, AN EDUCATION THEORY OF FAULT FOR AUTONOMOUS SYSTEMS (2017).

Solum, Lawrence B., Legal Personhood for Artificial Intelligences, 70 N.C. L. REV. 1231, 1244–48 (1992).

Stumpff, Andrew Morrison, The Law Is a Fractal: The Attempt to Anticipate Everything, 44 LOY. U. CHI. L.J. 649, 676 (2013).

Sullivan, Kathleen M,. Foreword: The Justices of Rules and Standards, 106 HARV. L. REV. 22, 66 (1992).

Thomas, W. Robert, How and Why Corporations Became (and Remain) Persons Under the Criminal Law, 45 FLA. ST. U. L. REV. 479, 504–14 (2018).

Thompson, Larry D., The Blameless Corporation, 47 AM. CRIM. L. REV. 1251, 1255 (2010).

Tremble, Catherine, Wild Westworld: Section 230 of the CDA and Social Networks' Use of Machine-Learning Algorithms, 86 FORDHAM L. REV. 825, 837 (2017).

Tutt, Andrew, An FDA for Algorithms, 69 ADMIN. L. REV. 83, 106 (2017).

Williams, Glanville, Vicarious Liability and the Master's Indemnity, 20 MOD. L. REV. 220, 230 (1957).

6

Artificial Intelligence as the End of Criminal Law? On the Algorithmic Transformation of Society

(https://doi.org/10.47907/livro2021_4c7)

Christoph Burchard[1]

Abstract:
Does Artificial Intelligence (AI) imply the end of criminal law and justice as we know it? This article submits that AI is a transformative technology that seemingly assumes and optimizes the rationalities of criminal law (the effective prevention of crime; the objective, neutral and coherent application of the law etc.), namely by replacing the

---

[1] Prof. Dr., LL.M. (NYU), Chair for Criminal Law and Criminal Procedure, International and European Criminal Law, Comparative Law and Legal Theory at the Goethe University Frankfurt am Main.

This contribution was first published in German with the title «Künstliche Intelligenz als Ende des Strafrechts? Zur algorithmischen Transformation der Gesellschaft», in Joerden Jan C./Schuhr Jan C., *Jahrbuch für Recht und Ethik. Zugleich Gedächtnisschrift für Joachim Hruschka*, Berlin: Duncker & Humblot, 2019, p. 527-555, and later in Italian «L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società», Rivista Italiana di Diritto e Procedura Penale, 4/62 (2019) p. 1909-1942. I would like to thank the editors and the publisher for letting me now publish an English translation in this volume.

The English translation has been curated by Mr. Johannes Abrell. To him goes my utmost gratitude.

This contribution was published as part of the research project "The Normative Order of Artificial Intelligence | NO:AI", which I am conducting as Principal Investigator at the Research Centre "The Formation of Normative Orders".

counterfactual guarantees of the law with the factual guarantees of technology. As a consequence, AI must not be trivialized by criminal law theory. Likewise, it is not enough to subversively criticize the current weaknesses of AI (e.g. vis-à-vis the "bias in, bias out" problem). Rather, criminal law theory should draw on the highflying promises of AI to reflect upon the foundational premises of criminal law. For a criminal law that is mostly a governance tool in the administrative and/or welfare state, AI applications promise the culmination of the law's very objectives (like the effective inhibition and prevention of crime, e.g. by means of predictive policing; or the political determination of fuzzy sentencing rationales in sentencing algorithms that ensure equal sentences for comparable crimes). For a criminal law, however, that protects liberal freedoms and rests on inter-personal trust, AI may well lead to the passing of the law's very ideals (e.g. of the presumption of innocence, which can no longer be upheld once everyone, ordinary citizens and judges alike, is deemed a possible risk). The question about "AI as the end of criminal law?" thus eventually raises the two-pronged question "Which criminal law for which society?". Indeed, what is the status of freedom (especially in a surveillance society needed to power Big Data driven algorithms), trust (especially under the zero trust paradigm that underlies many risk assessment algorithms) and future (especially when algorithms make predictions based on past data) once AI enters into the administration of criminal justice? These are the questions, or so I respectfully submit, that criminal law theory needs to address today in order to come up with a criminal law that is both (for pragmatic reasons) open to technology as well as (for humane reasons) sensible. In all of this, we must take to heart Joachim Hruschka's great legacy and remain intellectually honest.

Key-words: criminal law; artificial intelligence; predictive society; big data policing; predictive policing; risk assessment algorithms.

Question: "Can you imagine smart machines powered by AI being used to support judicial decision-making in the future?" – Answer: "This is already a reality. And it puts a very significant burden on the judiciary and how it functions." – Question to and answer by U.S. Supreme Court Chief Justice John Glover Roberts Jr. in April 2017[2]

INTRODUCTION

Shortly before I received the honorable invitation to contribute to this important commemorative volume, I had just examined the essay by *Byrd/Hruschka* on Kant's theory of crime in a seminar on "Criminal Law between Morality and Politics".[3] A truly groundbreaking essay, which removed Kant from the (for many unappealing) wing of absolute theories of crime. This marked the end of a journey for me. I first encountered *Joachim Hruschka's* groundbreaking oeuvre while working on my doctoral thesis. I can still remember vividly how impressed I was – then[4] and ever since – by his methodically and jurisprudentially reflected, equally logical-analytical and in the best sense enlightened approach to dogmatics. Therefore, it is a real honor for me, in memory of *Joachim Hruschka*, to be able to offer some initial thoughts today on recent developments, precisely on how so-called artificial intelligence "threatens" to change our society and our criminal law. This choice of topic may seem surprising at first, since there seem to be only a few direct points of reference to *Joachim Hruschka's* work. However, as I would like to show, the use of artificial intelligence poses fundamental challenges to our social and criminal legal order. What is needed no less than a reflection on the normative foundations of our criminal law, is a criminal law suited to a society, which is to be explained,

---

[2] Reported by Liptak, Adam, «Sent to Prison by a Software Program's secret Algorithms», *New York Times*, May 1, 2017, p. A22.

[3] Byrd, Sharon; Hruschka, Joachim, «Kant zu Strafrecht und Strafe im Rechtsstaat», Juristen Zeitung, 20/62 (2007) p. 957.

[4] I used the following titles, inter alia, in my doctoral thesis: Hruschka, Joachim, «Die Herbeiführung eines Erfolges durch einen von zwei Akten bei eindeutigen und bei mehrdeutigen Tatsachenfeststellungen», Juristische Schulung (1982) p. 317; – *Strafrecht nach logisch-analytischer Methode*, Berlin/New York: De Gruyter, 1988; – «Der Standard-Fall der aberratio ictus und verwandte Fallkonstellationen», Juristen Zeitung, 10/46 (1991) p. 488.

understood and evaluated anew. And in this regard, the virtual conversation with *Joachim Hruschka* needs to be continued in confrontation with his philosophical, not solely alleged,[5] but always justified insights, in order to be able to critically question today's developments. Less courageously than *Joachim Hruschka*, who once demanded no less than a rethinking of criminal law,[6] I advocate here to pass a reckoning on the society that brings forth or is to bring forth, our criminal law. To this end I must – as *Joachim Hruschka* recognized a long time ago[7] open the criminal law again to the social and political theory.

***

"Artificial intelligence (AI) and criminal law" is *not* science fiction. In light of the "digital revolution" permeating all areas of life, neither law in general nor criminal law in particular can escape the influences of AI.[8] As far as AI is already taken into consideration by criminal law scholars, it is traditionally (especially in Germany) considered as an area potentially in need of regulation[9] and also capable of regulation.[10]

This contribution is advocating for a *supplementing of perspective*.[11] In other words, to focus on AI (or more precisely its social practice) as a medium of contemporary social and (criminal) legal

---

[5] Decidedly against this Hruschka, Joachim, «Kann und sollte die Strafrechtswissenschaft systematisch sein?», Juristen Zeitung, 1/40 (1985) p. 10.

[6] Hruschka, Joachim, «Das Strafrecht neu durchdenken!», Goltdammer's Archiv (1981) p. 237.

[7] Id., p. 249. Also relevant to social criminal law theory Hruschka, Joachim, «Utilitarismus in der Variante von Peter Singer», Juristen Zeitung, 6/56 (2001) p. 261.

[8] See for instance the provocative questioning of Schwintowski, Hans-Peter, «Wird Recht durch Robotik und künstliche Intelligenz überflüssig?», Neue Juristische Online-Zeitschrift, 42 (2018) p. 1601.

[9] See in general Meyer, Stephan, «Künstliche Intelligenz und die Rolle des Rechts für Innovation», Zeitschrift für Rechtspolitik, 8/51 (2018) p. 233 for the relevant considerations to be made in this regard, including the need to maintain technological, innovative strength in international (economic and scientific) competition.

[10] The *agency* question is often prominent in this context, *i.e.*, who is responsible if damage is caused by the use of an AI system, especially when self-learning AI is used. On this topic in detail see Hilgendorf, Eric, «Autonome Systeme, künstliche Intelligenz und Roboter», in Barton, Stephan et al., *Festschrift für Thomas Fischer*, Munich: C.H. Beck, 2018, p. 111 ff.

[11] In general on this topic see Balkin, Jack B., «The Path of Robotics Law», California Law Review Circuit, 6 (2015) p. 45.

transformations.[12] The question should therefore be asked: How does AI already change our social and criminal legal system?[13] Does AI even mean, in exaggerated terms, the end of criminal law?[14] Be it either in the sense of a dying death of its fundamental principles (such as the presumption of innocence) or a crowning culmination of its fundamental goals (such as the protection of legal interests)?[15]

In order to answer these questions, it is necessary to examine the justification narratives more closely that are already being used today for the introduction of AI into criminal law; to analyze more closely the normative conceptions of order that lie behind these narratives and that are also concealed by them;[16] and in doing so, to highlight the power-political and ideological openness of these conceptions of order. In the spirit of this program, the promises of AI will first be illuminated with a view to current developments (see I. and II. below),

---

[12] AI can therefore be flagged as a potentially "*transformative technology*" to establish a dialectical connection between technological, societal, and legal change. On this topic in general see Fateh-Moghadam, Bijan, «Selbstbestimmung im biotechnischen Zeitalter», Basler Juristische Mitteilungen, 5 (2018) p. 205 (in particular p. 209 ff.). – By focusing on the social practice of AI as a medium of social transformations, we on the one hand oppose an essentialization of AI (see I. below) and on the other hand clarify that AI itself does not have a normative effect, but nevertheless conveys programmed ideas of order.

[13] The question insinuates a causality which, on closer inspection, can be resolved as a dialectical process in which social etc. developments promote the development and use of certain AI systems, which in turn reinforce the first-mentioned developments etc. AI hereby "naturally" builds on the general mechanization of real life. For example, the use of electronic ankle bracelets, especially against so-called endangerers, could be seen as a glimpse of a partial surveillance society, which always deprives certain groups of people (precisely these so-called endangerers) of the counterfactual trust in their lawfulness. This can then be reinforced by AI with a general *zero trust* paradigm. In this respect see II. below.

[14] This question is meant to shake things up and not to express a *fin de siècle*. The point is: how are we to deal with the polyvalences of today's developments? For a similar play on words see Hildebrandt, Mireille, *Smart Technologies and the End(s) of Law*, Cheltenham: Edward Elgar Publishing, 2015.

[15] My questions are *not* to be understood as speculative science fiction, so that general dystopias are not considered here.

[16] For an introduction to this conceptual apparatus coined at the Cluster of Excellence "Die Herausbildung normativer Ordnungen" see Forst, Rainer; Günther, Klaus, «Die Herausbildung normativer Ordnungen: zur Idee eines interdisziplinären Forschungsprogramms», in Forst, Rainer; Günther, Klaus, *Die Herausbildung normativer Ordnungen*, Frankfurt a.M.: Campus, 2011, p. 11 (in particular p. 15 f.).

in order to then reflect on how a criminal law theory, that opens up critical perspectives,[17] should deal with them (see III. and IV. below).

Particular attention is paid in this respect to promises that supposedly smart or intelligent algorithms that regularly evaluate Big Data will provide more effective and efficient protection of legal interests and allow for a more neutral, objective, and coherent law enforcement than human decision makers. These promises correspond *prima facie* to those of criminal law. As the supposedly *ultima ratio* of the state, criminal law also promises a particularly thorough protection of legal interests. Moreover, criminal law demands for its impartial and unbiased as well as consistent application. The difference between the two is that AI promises technological facticity, while criminal law – like law in general – can only profess counterfactual guarantees. Whether AI means the end of criminal law is in consequence synonymous with the question, which criminal law is meant. A liberal criminal law of freedom, which is based on interpersonal trust and does not manage people solely as controllable and constantly assessable potential risks (in other words, as endangerers), is fundamentally called into question by AI; a liberal (criminal) law theory would therefore have to seek the forward defense of the counterfactual of (criminal) law against the factual of AI. For a welfare-state security criminal law that sees itself as an instrument of social control or governance of social interactions, AI on the other hand enables a crowning culmination of its rationality.

I. AI AS A NON-ESSENTIALIST CONSTRUCT

AI is not an unambiguous term. The exact definition of the term is being fought over on all sides.[18] One often encounters what is

---

[17] Criticism is understood here axiomatically as a practice of justified doubt. The normative program lies in the opening of possibilities for critical, doubting inquiries, not in the development of a normative program from which criticism can be exercised. The latter would be reserved for a critical theory of criminal law, which cannot be developed here. A return to a rational concept of critique is also demanded by Hruschka, Joachim, *Strafrecht nach logisch-analytischer Methode*, Berlin/New York: De Gruyter, 1988, p. XI.

[18] Of fundamental importance are debates about whether the name AI even accurately represents the current state of research and development (e.g., if and because

essentially a structural essentialism that seeks to fathom the proprietary properties of intelligence in general and artificial intelligence in particular.[19] This may also explain why profound discussions about the criminal responsibility of intelligent and self-aware machines of the future[20] are already being held.[21]

Here, nevertheless, a non-essentialist understanding of AI is advocated in order to derive insights from the perspective of an observer and make these of use for a modern and critical theory of criminal law, that thinks in the categories of the participant perspective. Thus, it is not about what constitutes or should constitute the essence of

---

today's "AI" systems do not go beyond classical machine learning and time-honored pattern recognition); whether AI is really "intelligent" (e.g., if and because today's "AI" systems cannot provide transfer services); and whether the German qualification "künstlich" is correct (e.g., if and because the qualification "maschinelle Intelligenz" is meant to represent the properties of algorithms more correctly, or the "artificial" is pejorative in the romantic sense). On this topic see Herberger, Maximilian, «„Künstliche Intelligenz" und Recht», Neue Juristische Wochenschrift, 39/71 (2018) p. 2825. – Cf. further (the only at first seemingly outdated work of) Weizenbaum, Joseph, *Die Macht der Computer und die Ohnmacht der Vernunft*, Frankfurt a.M.: Suhrkamp, 1977, p. 268 ff.

[19] This is how for instance Ertel, Wolfgang, *Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung*, Wiesbaden: Springer VS, 2016, p. 1 ff. views the questions "What is intelligence?", "How can intelligence be measured?", or "How does our brain work?" as significant for the understanding of AI. It is further alleged that for computer scientists and in particular engineers, the question "about the intelligent machine that behaves like a human being, that shows intelligent behavior," is decisive.

[20] This is reflected in an essentialist and anthropocentric understanding of criminal law, which needs to be questioned in the course of the worldwide triumph of associational and corporate criminal law. It does not seem far-fetched (and nothing further is put up for discussion here) to want to justify the criminal liability of AI systems strictly functionally, *i.e.* independent of how intelligent an AI system now is and whether it is aware of itself and whether it can thus be ascribed original human characteristics. The following points, in analogy to corporate criminal liability, could be mentioned functionally in favor of an AI criminal liability: (1) Indirectly, the owners of a deficient AI system should be targeted (e.g., if the shutdown of the AI system is ordered) (2) The recourse to responsible persons who have developed a deficient AI system or brought it to the market shall be cut off (e.g. because and if there are innumerable persons "behind" an AI system, from programmers to company managers of any hierarchical level, so that an individualization of responsibility would not be feasible). (3) Or, by attributing criminal responsibility, feelings of resentment and indignation are to be expressed in an institutionalized form (I thank *Boris Burghardt* for pointing out this aspect).

[21] For further detail see Gaede, Karsten, *Recht und Strafen für Roboter?*, Baden-Baden: Nomos, 2019 with further references.

(artificial) intelligence. Rather, the starting point is to ask how AI (especially through corresponding justification narratives) is socially constructed, represented and received; how AI is embedded in certain social relations and changes them; and which relations of domination and power are expressed in AI and stabilized, mystified, transformed or produced by it.[22] In other words, it is a matter of assigning meanings that originate in the social reality of life and have an effect on it. In this process, normative conceptions of order (including ideologies) which are to be determined in each case are promoted and specific conflicts are either directed to the center of attention or cast into the background. Seen in this light, AI is a normatively open and malleable, and simultaneously political construct. It is exactly the ambiguity and power-political openness of the term AI that allows interested actors to instrumentalize it for their own – political, economic, etc. – purposes. AI is a particularly powerful name, because it attests to the general cognitive ability of an IT system to "understand and learn well, and to form judgements and opinions based on reason" in purely linguistic terms – i.e., independently of its "real" intelligence.[23]

In this context, AI should *not* be seen as a neutral technology or simply as an information technology innovation. Rather, AI is directly linked to the basic principles of human sociality (freedom, tolerance, law, etc.), traces back to them and transforms them. Therefore, it is important to take the promises, hopes and fears associated with the term AI at any given time seriously in order to be able to subject them to a critical reflection, as they are fueled and spread with a varying urgency by the most diverse actors in business, politics, science, etc.

As an example, let us refer to the conventional image of AI that has been cemented in our minds, thanks to *Hollywood*.[24] Here, AI

---

[22] Similarly Balkin, Jack B., «The Path of Robotics Law», California Law Review Circuit, 6 (2015) p. 59; Weizenbaum, Joseph, *Die Macht der Computer und die Ohnmacht der Vernunft*, Frankfurt a.M.: Suhrkamp, 1977, p. 268 ff. See also Mau, Steffen, *Das metrische Wir: Über die Quantifizierung des Sozialen*, Berlin: Suhrkamp, 2017.

[23] Intelligence is understood here in terms of the third definition as provided by the online Cambridge dictionary.

[24] For an overview on this topic see Xanke, Lisa; Bärenz, Elisabeth, «Künstliche Intelligenz in Literatur und Film – Fiktion oder Realität?», Journal of New Frontiers in Spatial Concepts, 4 (2012) p. 36; Irsigler, Ingo; Orth, Dominik, «Zwischen Menschwerdung und Weltherrschaft: Künstliche Intelligenz im Film», Aus Politik und Zeitgeschichte, 6-8/68 (2018) p. 39.

stands for autonomous robots (one thinks dystopic of "Terminator" or more overtly "I, Robot"), intelligent androids (one thinks of Lieutenant Commander Data in "Star Trek: The Next Generation"), and self-aware supercomputers (think of HAL 9000 in "2001: A Space Odyssey" or Central in "Star Trek: Discovery"). In this respect, AI appears (which is dramaturgically understandable) almost exclusively as so-called *strong AI*, which strives for the same general intelligence as humans or has already attained it, if not surpassed it.[25] The main focus of the discussion is then on both the *conditio humana* (whether and how humans can integrate into a machine world)[26] and the *conditio automata* (whether and how intelligent machines can integrate into a human society).[27] Since none of the systems in existence today falls under the category of strong AI,[28] these discussions admittedly do not yet have any immediate practical legal significance. This may be one reason why "AI and criminal law" has been an orchid topic at best until recently, and why the inherent characteristics of strong artificial intelligence continue to be fought over in an essentialist manner.

However, we may not forget that we are currently becoming the witnesses of another AI revolution.[29] This revolution is based, in substance, on so-called *weak AI*, which is optimized for solving problems for specific applications, is based on known methods of mathematics and computer science, and does not acquire a deeper – or actual – understanding of problem solving.[30] And in which – in contrast to Hollywood – no robots or androids are used.

---

[25] For the differences between weak and strong AI see the "classic" Searle, John R., «Minds, Brains, and Programs», Behavioral and Brain Sciences, 3/3 (1980) p. 417.

[26] As a literary example see: McEwan, Ian, *Machines like Me*, London: Penguin, 2019.

[27] As a cinematical example see: *Ex Machina*, 2015.

[28] If one follows the conceptual distinction between weak and strong AI as developed by Searle, John R., «Minds, Brains, and Programs», Behavioral and Brain Sciences, 3/3 (1980) p. 417, this distinction is in its present use still thoroughly inconsistent. The central vanishing point of a strong AI is seen in the equivalent to human abilities. On this see Ramge, Thomas, *Mensch und Maschine: Wie künstliche Intelligenz und Roboter unser Leben verändern*, Stuttgart: Reclam, 2018, p. 19.

[29] Essentially as here, the popular science work by Fry, Hannah, *Hello World: Was Algorithmen können und wie sie unser Leben verändern*, München: C.H. Beck, 2018.

[30] In favor of this, despite all opposition, Ramge, Thomas, *Mensch und Maschine: Wie künstliche Intelligenz und Roboter unser Leben verändern*, Stuttgart: Reclam, 2018, p. 19.

The fact that these systems are also listed here as AI is a consequence of the methodological approach we have just adopted. Due in part to the corresponding efforts of industry,[31] we in politics and society are increasingly detaching ourselves from AI à la Hollywood and understand AI to mean information technology systems that are "sold" to us as solutions to real-life problems – and that we also "buy" for this purpose. Indicative of this is the recent concept paper of the Council of Europe's "European Committee on Crime Problems (CDPC)" on "AI and Criminal Liability". There, self-driving cars along with their self-learning algorithms are flagged as prime examples of AI as though this were self-evident.[32] The Council of Europe's "European Commission for the Efficiency of Justice (CEPEJ)" is playing the same tune by advocating the use of AI in the administration of justice in certain areas and under certain conditions.[33] All this is an expression of today's AI hype. This hype is carried by the multilaterally reproduced narrative that smart or intelligent algorithms are capable of mastering real-life problems that exceed human capabilities by means of information technology to the benefit of all; and in fact do so better, faster and more cost-effectively than human decision-makers (beginning with the safe control of cars and continuing over to the evaluation of all medical publications in support of disease diagnoses and therapy concepts up to the automation of legal services, so-called *legal tech*[34]).

This "new" meaning of AI is not neutral, but normative and also ideologically charged. One should, despite the ongoing struggle for interpretive dominance, be careful not to speak of "the" – all the while dominant – meaning of AI. Nonetheless, *Katz's* finding is convincing

---

[31] For an elaboration and critique see Zuboff, Shoshana, *Das Zeitalter des Überwachungskapitalismus*, Frankfurt a.M./New York: Campus, 2018.

[32] European Committee on Crime Problems (CDPC), "Artificial Intelligence and its Impact on CDPC Work: The case of automated driving", CDPC (2018) 14 – 14.09.2018, p. 6 *et seq.* [Date of consultation 14.12. 2021], Access: https://rm.coe.int/cdpc-2018-14-artificial-intelligence-and-criminal-law-project-2018-202/16808d6d09.

[33] CEPEJ, "European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment" v. 3 – 4.12.2018, in particular p. 64 *et seq.* [Date of consultation 14.12. 2021], Access: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.

[34] For overviews on this topic see Fries, Martin, «Automatische Rechtspflege», Rechtswissenschaft, 4/9 (2018) p. 414.

that "the "AI" label has been rebranded to promote a vision of world governance through big data."[35] This vision is often combined with a metaphysically and theologically charged hope of redemption. AI promises a self-learning and self-improving entity that promises transparency and predictability of the other as well as the self, and thus hyper-rational controllability of social interactions.[36] This narrative has particular social and political traction and explosive power, since it is generalized and permeates all areas of life (which already sets AI apart quantitatively from earlier particularistic "criminological" currents à la Lombroso). The narrative, moreover, is so generically construed that it becomes neoliberally,[37] rationally and scientifically,[38] liberally[39] as well as authoritatively[40] appropriable.[41] AI, or rather their fundamental normative postulates of order and narratives of justification, are in other words latently open to (power) politics. And in them, a new manifestation of the dialect of enlightenment is also revealed.

II. THE PROMISES OF AI FOR CRIMINAL JUSTICE

After having previously addressed the general promises made by AI, we will now address the specific promises made by AI for criminal law or criminal justice. Namely, the promise that crime can actually

---

[35] Katz, Yarden, «Manufacturing an Artificial Intelligence Revolution», SSRN (2017) p. 1, [Date of consultation 14.12.2021], Access: https://ssrn.com/abstract=3078224.

[36] In this regard, the stance of Nida-Rümelin, Julian; Weidenfeld, Natalie, *Digitaler Humanismus*, München: Piper, 2018, p. 44 *et seq.* is accurate.

[37] Critical in this regard Zuboff, Shoshana, *Das Zeitalter des Überwachungskapitalismus*, Frankfurt a.M./New York: Campus, 2018.

[38] Critical in this regard Nida-Rümelin, Julian; Weidenfeld, Natalie, *Digitaler Humanismus*, München: Piper, 2018, Introduction.

[39] As for instance Chiao, Vincent, «Predicting Proportionality: The Case for Algorithmic Sentencing», Criminal Justice Ethics, 3/37 (2018) p. 238 ff.

[40] For an overview on this – and in particular regarding China – see Mau, Steffen, *Das metrische Wir: Über die Quantifizierung des Sozialen*, Berlin: Suhrkamp, 2017, p. 9 ff.

[41] Theoretically, it can then be said that the justification narratives of AI (the first analytical point of reference) refer to underlying normative orders (the second analytical point of reference), which are themselves open to power politics or ideology (the third analytical point of reference).

be made impossible, or can in any case be drastically reduced, through the use of intelligent information technology (see 1. below); and that decision making in criminal justice can be exempt from human subjectivity and bias and therefore can "finally" be truly objective, neutral, and coherent (see 2. below).

1. Efficacy and Efficiency in the Inhibition of Crime

First, AI promises to make crime impossible, both directly and indirectly. Or more precisely: advocates promote that certain forms of crime are *eo ipso* no longer committable through the use of AI, or that the commission of certain forms of crime is *de facto* significantly reducible through AI-supported (sovereign, privatized, or internalized) enforcement and surveillance structures. This is illustrated by so-called Smart Contracts[42] and Predictive or Big Data Policing[43].

a) Smart Contracts

Smart contracts are qualified as "smart", i.e. as clever, witty and sophisticated contracts, on the basis of specific information technology requirements. They appear with the promise of being able to handle contractual interactions more effectively and efficiently. Smart contracts are implemented through computer programs that aim to algorithmically enable, verify, and enforce contractual rights and obligations without relying on third parties. The idea is to automate contract drafting and execution to the greatest possible extent, which is ideally "self-executing" and thus minimizes the transaction costs of traditional contract law. Through this, the legal system (including its representatives such as notaries or judges) but also private service providers who insure against payment or delivery defaults are to be made obsolete.

---

[42] Smart contracts are generally discussed in private rather than criminal law literature, so further references have been omitted here. Smart contracts are not necessarily to be subsumed under AI. However, since corresponding developments are becoming apparent, they are presented here.

[43] Big Data Policing describes a recent development, particularly in the USA, which has yet to be developed in Germany in terms of terminology. For this see Ferguson, Andrew G., *The Rise of Big Data Policing. Surveillance, Race, and the Future of Law Enforcement*, New York: New York University Press, 2017. See also the contributions in the Ohio State Journal Criminal Law 2018, p. 473 ff. on a "Round Table on Big Data and Criminal Law".

Free by the credo "code is law", the law as well as the application of the law is replaced by a corresponding IT infrastructure.

Smart contracts are based on "distributed ledger technology" such as the so-called blockchain, i.e. a database networked with the reality of life, which is de-centrally stored, verified and continuously updated. And in which, for example, is stored who has which goods or financial resources. The more comprehensive and accurate this database is – that is, the more data records (Big Data) it has at its disposal – the more precisely and securely the smart contract can and will be executed – at least that is the claim.

This is supported by the vision of a trust-free society in which contract and/or interaction partners no longer have to trust each other.[44] This is because – according to the justification narrative – they are given better, namely absolute or unchallengeable informational assurances (e.g. that the seller has the offered goods at his disposal and will hand them over and transfer ownership; and that the buyer is sufficiently liquid and will actually pay for the goods). This is accompanied by a shift in trust, away from interpersonal trust and towards trust in information technology systems (the database and programming).[45] The contract and interaction partners are basically considered as a risk, since it – indeed – cannot be ruled out that they will breach their word or the contract. To cope with this risk, a hand is laid on interpersonal trust as a counterfactually postulated, as it is simply a socially necessary, mechanism for reducing social complexity.[46] "Smart" databases and algorithms are supposed to ensure that people interacting socially rely on each other, because they know about one another, that the possibility of database- and algorithm-unfriendly behavior is taken away from them by "smart" algorithms and databases. One person relinquishes the possibility that they might behave in disconformity to the databases and algorithms, because, if and so that the other person

---

[44] In general see Palka, Silvia; Wittpah, Volker, «Vertrauen und Transparenz – Blockchain Technologie als digitaler Vertrauenskatalysator», Working Paper of the Institute for Innovation and Technology, 39 (2018).

[45] For more on this and in general see Wagner, Gerald, «Vertrauen in Technik», Zeitschrift für Soziologie, 2/23 (1994) p. 145.

[46] According to the classical definition of Luhmann, Niklas, *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität*, 5. Ed., Munich: UVK, 2014, p. 27 ff. (in particular p. 30).

is deprived of this possibility as well. To put it bluntly: *homo homini lupus est* is not only psychologically displaced (trust-based interaction), but also averted from the outset in terms of information technology (so-called *zero trust* or *in tech we trust* interaction).

In terms of criminal law, this is intended to make it impossible to commit fraud in the case of exchange contracts, when leaving sovereigns and the privatizing of crime prevention out of the equation.[47] This can be seen in the "grail scripture" of the original blockchain movement, with and in which the Bitcoin idea was invented and explained. Here it is stated that:

> "Commerce on the Internet has come to rely almost exclusively on financial institutions serving as trusted third parties to process electronic payments. While the system works well enough for most transactions, it still suffers from the inherent weaknesses of the trust based model. Completely non-reversible transactions are not really possible, since financial institutions cannot avoid mediating disputes. The cost of mediation increases transaction costs, limiting the minimum practical transaction size and cutting off the possibility for small casual transactions, and there is a broader cost in the loss of ability to make non-reversible payments for nonreversible services. With the possibility of reversal, the need for trust spreads. Merchants must be wary of their customers, hassling them for more information than they would otherwise need. A certain percentage of fraud is accepted as unavoidable. These costs and payment uncertainties can be avoided in person by using physical currency, but no mechanism exists to make payments over a communications channel without a

[47] Further examples of such a privatization of crime prevention through AI are provided, for instance, by AI-supported "criminal compliance" systems (i.e., Digitial Compliance tools), which promise e.g., full monitoring of internal company communications with the aim of "flagging" suspicious interactions and thus subjecting them to further scrutiny. For an overview on this see Schemmel, Alexander; Dietzen, Alexandra, «"Effective Corporate Governance" by Legal Tech & Digital Compliance», in Breidenbach, Stephan; Glatz, Florian, *Rechtshandbuch Legal Tech*, Munich: C.H. Beck, 2018, p. 137 – In Japan, as can be seen from press reports, there is also controversy about whether AI should be used to prevent shoplifting. For this purpose, customers can be monitored algorithmically in order to be able to make predictions from the analysis of their body language, as to whether they are planning to shoplift. On this see Lewis, Nell, «Should AI be used to catch shoplifters?», *CNN Business* (2019) [Date of consultation 14.12.2021] Access: https://edition.cnn.com/2019/04/18/business/ai-vaak-shoplifting/index.html.

> trusted party. What is needed is an electronic payment system based on cryptographic proof instead of trust, allowing any two willing parties to transact directly with each other without the need for a trusted third party. Transactions that are computationally impractical to reverse would protect sellers from fraud, and routine escrow mechanisms could easily be implemented to protect buyers."[48]

b) Predictive/Big Data Policing

Predictive policing and *big data policing*, which originated in the United States and are now also taking hold in other Western countries, including Germany,[49] are also based on corresponding narratives. Here, as there, it is about – according to the presentation by producers and distributors – intelligent algorithms that provide resource-scarce authorities with powerful instruments for the prospective prevention of crimes.[50] *Brantingham*, a scientific pioneer of predictive policing, defines it according to the following three-step process:

> "(1) data of one more type are ingested; (2) algorithmic methods use ingested data to forecast the occurrence of crime in some domain of interest; and (3) police use forecasts to inform strategic and tactical decisions in the field. A primary goal of predictive policing is to reduce uncertainty so that police can approach the allocation of resources in an optimal manner. The theory is that an optimal allocation of police resources has a better chance at disrupting opportunities for crime before they happen."[51]

Big Data Policing is mainly set up accordingly. *Big Data Policing* also revolves around algorithmic predictions that are intended to

---

[48] Nakamoto, Satoshi (a pseudonym!), «Bitcoin: A Peer-to-Peer Electronic Cash System» (2009) p. 1, [Date of consultation 14. 12. 2021], Access: https://bitcoin.org/bitcoin.pdf.

[49] For an overview on corresponding (pilot) projects in Germany see Rademacher, Timo, «Predictive Policing im deutschen Polizeirecht», Archiv des öffentlichen Rechts, 3/142 (2017) p. 369.

[50] For a similar description see the coalition agreement between CDU Hessen and BÜNDNIS90/DIE GRÜNEN Hessen for the 20th legislative term, line 2535 ff., where it can be read: "New tools such as specialized data processing systems that pool and analyze existing information from police databases can be of great benefit in addressing current policing challenges.".

[51] Brantingham, Jeffrey, «The Logic of Data Bias and Its Impact on Place-Based Predictive Policing», Ohio State Journal of Criminal Law, 2/15 (2018) p. 473.

enable authorities to prevent crimes as effectively and efficiently as possible – in this logic, "ideally" and with regard to specific persons.[52] The technological difference to Predictive Policing lies in the quantity and quality of data sets that go into Big Data Policing. As the name suggests, it processes large, seemingly disjointed data sets that traditional analysis tools (including the human brain) were no match for.

Behind the scenes, big data policing is likely to be more far reaching than simple predictive policing, since it is directed not only at public authorities, but also at citizens. As *Brennan-Marquez* has aptly noted, Big Data Policing generates "a social order – a surveillance society – in which people constantly monitor and curate the data-trails they leave behind in everyday life."[53] In other words: the more (data) intensive predictions on behavior and crime inhibition turn out to be, the more the internalization of this crime inhibition project is promoted and the more the external (soverign/public) and internal (private) law enforcement complement each other. In this respect, a return to *Foucault's* panoptism seems of interest. Because, by committing himself to the project of technological crime inhibition through risk surveillance of others, the individual does so at the price of being surveilled themselves, accepts this and thus becomes the enforcer of the power structures that lie beneath this surveillance project.[54]

The fact that the algorithms, on which predictive or big data policing is based – as described by the relevant nomenclature – make "decisions"[55] (especially because they establish correlations between seemingly unrelated data sets), is considered sufficient by quite a few voices, especially in the USA.[56] This is considered to be the case, even

---

[52] Brennan-Marquez, Kiel, «Big Data Policing and the Redistribution of Anxiety», Ohio State Journal of Criminal Law, 2/15 (2018) p. 487.

[53] Brennan-Marquez, Kiel, «Big Data Policing and the Redistribution of Anxiety», Ohio State Journal of Criminal Law, 2/15 (2018) p. 487.

[54] Euphemistically, this is called self-documentation, the other side of which is *Foucault's* panopticon of the self and a moment of the exercise of power. On this see Mau, Steffen, *Das metrische Wir: Über die Quantifizierung des Sozialen*, Berlin: Suhrkamp, 2017, p. 249 ff. – For more see Han, Byung-Chul, *Psychopolitik: Neoliberalismus und die neuen Machttechniken*, Frankfurt a.M.: Fischer, 2015, p. 84.

[55] Anthropomorphisms dominate the social construction of AI today, even as they eclipse the technical and algorithmic idiosyncrasies of AI. It is at least misleading to think that AI systems make decisions, since they can actually only generate output.

[56] As here Henderson, Stephen E., «A Few Criminal Justice Big Data Rules», Ohio State Journal of Criminal Law, 2/15 (2018) p. 527.

if these decisions are not comprehensible or explainable.[57] Simply put, it should be possible to keep the algorithms – in contrast to traditional prediction models – deliberately "atheoretical".[58]

In other words, and with the corresponding *termini technici*: so-called "*opaque AI*" is considered acceptable and "*explainable AI*"[59] is considered negligible.[60] It seems reasonable to see in this an increased faith in technology and AI, which in turn is fueled by a loss of confidence in human analytical and decision-making capabilities. In any case, this perspective of things demonstrates the vision of an AI-supported organization and control of human sociality and interactivity. This vision is particularly powerful in the context of criminal law. After all, who would want to deny that effective and efficient crime prevention is desirable from the point of view of society as a whole, as well as from the individuals (e.g., the notorious victim's) point of view?

2. Objectivity, Neutrality and Coherence in the Application of Criminal Law

Given that we have previously spoken of the promise that AI is capable of effectively and efficiently minimizing the possibilities of committing crimes, we should now go into the further promise that the introduction of AI into (criminal) legal decision-making will make it possible to vouch for its objectivity, neutrality and coherence. The idea that decisions on the imposition of pre-trial detention due to the risk of repetition, on early release from prison due to a positive social prognosis or on sentencing could be made on the basis of *algorithmic risk assessment* may sound outrageous in the German Republic of Judges, due to the seemingly associated encroachment on judicial independence. In

---

[57] Although this is strongly – and in my opinion rightly – disputed by computer scientists or legal scholars. Critically for instance Liu, Han-Wei;Lin, Ching-Fu; Chen, Yu-Jie, «Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability», International Journal of Law and Information Technology, 2/27 (2019).

[58] As literally stated by Berk, Richard; Hyatt, Jordan, «Machine Learning Forecasts of Risk to Inform Sentencing Decisions», Federal Sentencing Reporter, 4/27 (2015) p. 223.

[59] Which can also be conceived and designed as "self-explanatory AI".

[60] As stated for instance by Henderson, Stephen E., «A Few Criminal Justice Big Data Rules», Ohio State Journal of Criminal Law, 2/15 (2018) p. 527.

the USA, however, this is already common practice, approved by the highest courts.

This is exemplified by *State v. Loomis*, a decision of the Supreme Court of Wisconsin.[61] The accused *Eric Loomis*, who had multiple prior criminal convictions, was suspected of being the driver of a "drive-by-shooting". Subsequently, he pleaded guilty to eluding an officer and did not contest the charge of operating a vehicle without its owner's consent. For this he was sentenced to six years in prison. This draconian sentencing, as the sentencing court openly acknowledged,[62] was due in part to the fact that *Loomis* had been diagnosed as having an abysmal social record and a high recidivism rate. And namely by COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), a proprietary (vulgo secret) algorithm developed and distributed by Northpointe, Inc. COMPAS calculated the accused's pre-trial and general risk of recidivism as well as his risk to the community based on a complex analysis of a 137-item questionnaire and the accused's public criminal record. *Loomis* appealed the decision on the grounds that it violated his right to due process. In particular, he challenged that he could not review the algorithmic processes because they were protected as trade secrets; that there was no individual penalty assessment because COMPAS worked with generalizing group data; and that the algorithm also processed the gender of the person(s) being assessed and thus an inadmissible variable, as it was gender discriminatory. – The Supreme Court of Wisconsin rejected the appeal, and the U.S. Supreme Court also ultimately did not accept the case for decision, after previously asking the U.S. federal government to comment on it. Of significance was the decision, that individuals should not enjoy a "right to explanation" of an algorithmic risk prediction as long as they can oversee its input and are informed of its output. Access to the throughput, e.g. why individual data blocks are weighted and how, was thus in principle legally denied to *Loomis*. The use of general group data and the inclusion of gender in the algorithmic risk prediction were also not objected to because this improved its accuracy and did not have a discriminatory objective. Nevertheless, the Supreme Court

---

[61] *State v. Loomis,* 881 N.W.2d 749 (Wis. 2016) 754 (US).

[62] For further details, in particular on the (in Wisconsin obviously procedurally admissible) use of a broader crime suspicion, which was added to the files, cf. the summary of the process in *State v. Loomis,* 881 N.W.2d 749 (Wis. 2016) 754 (US).

of Wisconsin was anxious to clarify that a judge may only take into account a corresponding algorithmic risk forecast, but may not regard it as binding.

State v. Loomis is received controversially and mostly negatively in the academic literature – inside and outside the USA.[63] However, this should not obscure the fact that algorithm-assisted legal decision-making is now a practice sanctioned by the highest courts in the USA. The causes (but note: not the reasons) for this are numerous. If one detaches oneself from the jurisdiction-specific analysis that is actually indicated,[64] the recurring motives of today's AI justification narrative can be found quickly. According to this narrative, the use of intelligent algorithms should make criminal justice decision-making processes more effective and efficient. More accurate risk predictions should be made and freed up resources should be able to be used elsewhere, e.g. in rehabilitation programs.[65] And while detention and sentencing decisions are dismissed as gut decisions, and even seen as *black art*,[66] algorithmic-based decision making is praised for apparently being able to minimizing the influence of biases, prejudices, and idiosyncrasies.[67]

---

[63] Cf. for instance Beriain, Iñigo De Miguel, «Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling», Law, Probability and Risk, 1/17 (2018) p. 45; Deskus, Cassie, «Fifth Amendment Limitations on Criminal Algorithmic Decision-Making», NYU Journal of Legislation and Public Policy, 1/21 (2018) p. 237; Liu, Han-Wei; Lin, Ching-Fu; Chen, Yu-Jie, «Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability», International Journal of Law and Information Technology, 2/27 (2019) p. 122 ff.; Huq, Aziz Z., «Racial Equity in Algorithmic Criminal Justice», Duke Law Journal, 6/68 (2019) p. 1081. – Cf. also Ostermeier, Lars, «Der Staat in der prognostischen Sicherheitsgesellschaft», in Puschke, Jens; Singelnstein, Tobias, *Der Staat und die Sicherheitsgesellschaft*, Wiesbaden: Springer VS, 2017, p. 103.

[64] To give just a few key words: A deep-seated racism, which is still fueled today at the highest political level, finds expression not least in the phenomenon of so-called "mass incarceration" and causes a – sad but understandable – feeling of resignation that the U.S. criminal justice system can no longer be "saved" by conventional means.

[65] See for instance Botnick, Claire, «Evidence-based Practice and Sentencing in State Courts: A Critique of the Missouri System», Washington University Journal of Law & Policy, 1/49 (2015) p. 166.

[66] As strikingly stated by Chiao, Vincent, «Predicting Proportionality: The Case for Algorithmic Sentencing», Criminal Justice Ethics, 3/37 (2018) p. 238.

[67] For a critical approach to this see Martini, Mario; Nink, David, «Wenn Maschinen entscheiden … – vollautomatisierte Verwaltungsverfahren und der Persönlichkeitsschutz», Neue Zeitschrift für Verwaltungsrecht – Extra, 10/36 (2017) p. 9.

Not the law (due to its postulated inherent properties) or the legal staff (due to its formation), but algorithms, in other words, are supposed to guarantee (and in the U.S.: save) the objectivity, neutrality and coherence of the application of law. Here we once again encounter the shift in trust already noted above, away from trust in people and toward trust in high technology. It is exactly this shift of trust that occurs for almost all humans, since, here the practitioners of the law are also perceived as a risk in the matter, namely as a risk to the objectivity, neutrality and coherence of the application of the law.[68]

III. THE PROMISES OF AI FROM THE PERSPECTIVE OF CRIMINAL LAW THEORY

The foregoing illustrates that criminal law theory can no longer ignore the promises of AI. These have developed too much social traction and explosive power to be ignored, and are in the process of becoming effectively entrenched in the criminal justice system. A trivialization of AI (see 1. below) as well as a subversive doubting of its promises (see 2. below) therefore does not seem to be a viable way to face the current challenges.

1. Criminal Sociological Trivialization?

A first, almost involuntary reaction to the promises described under I. and II. is to downplay them in their effects and potency and therefore to meet them with empathetic forbearance. Especially because many of the developments mentioned in the foregoing originate in the USA (or even in China) and therefore "cannot" develop any significance for Europe and Germany. Moreover, AI is not infrequently downgraded to a "normal" technological innovation that may be capable of transforming the criminal justice system to the usual extent and according to the usual pattern, but not of revolutionizing it (comparable, for example, to the introduction of modern passenger cars,

---

[68] Similar developments seem to be taking place in China, although there the center of discussion may be the control exercised by apparently dependent judges. Instructive on this matter: Meng, Yu; Guodong, Du, «Why Are Chinese Courts Turning to AI?», *The Diplomat* (2019) [Date of consultation 14.12. 2021], Access: https://thediplomat.com/2019/01/why-are-chinese-courts-turning-to-ai/.

which made an actual criminal traffic law necessary). All that can then be expected (and, depending on one's point of view, feared) is that the familiar developments of modern criminal law will continue, such as the forward shifting of criminal liability or the protection of collective legal interests. As a consequence, a fundamental questioning of criminal law could not be envisaged in terms of criminal law theory.

It could seemingly (!) be said, for instance: Even if *arguendo* the promises of smart contracts were to be taken at face value, this would merely lead to shifts and adjustments in where, when and how criminal (fraudulent) energy comes to bear.[69] In particular, an increase in cybercrime could be expected. Instead of directly deceiving a buyer into believing that one can dispose of the item to be sold as the seller, the latter will probably manipulate the database ("blockchain") that verifies the power of disposal and thus carry out this deception indirectly. Furthermore, active attacks on the computer programs contouring and executing a smart contract as well as the exploitation of their vulnerabilities are conceivable.[70] If such developments create gaps in criminal liability, it would be up to the legislator to close them. In doing so, the legislator will rely on the use of preemptive offenses aimed at protecting collective rights (such as the integrity and correctness of decentrally organized databases). – Furthermore, with regard to predictive or big data policing, it could be argued that algorithmic crime predictions only work, if at all, for specific crime areas (such as burglary and drug-related crime, which can be prevented by locally targeted police patrols).[71] The majority of the "fight" against crime would therefore have to be carried out by classical means.

---

[69] Parallels could be drawn here from Durkheim, Emile, *Die Regeln der soziologischen Methode*, Frankfurt a.M.: Suhrkamp, 1984, p. 86 and 156.

[70] An illustrative example of this is provided by the so-called DAO-Hack. On this see Heckmann, Jörn, «DAO- Hack: smart contracts auf dem rechtlichen Prüfstand», Computer und Recht, 9 (2016) R99.

[71] On the spatial dimension of predictive policing see for instance Straube, Till; Belina, Bernd, «Policing the Smart City: Eine Taxonomie polizeilicher Prognoseprogramme», in Bauriedl, Sybille; Strüver, Anke, *Smart City – Kritische Perspektiven auf die Digitalisierung in Städten*, Bielefeld: transcript Verlag, 2018, p. 223. Predictive policing is not limited to spatial crime forecasting, however. The – probably highly unfruitful – analysis of airline passenger data also falls under the heading of predictive policing. On this see Süddeutsche Zeitung-Online 24.04. 2019, "Überwachung von Flugpassagieren liefert Fehler über Fehler".

In addition, predictive or big data policing could give rise to new forms of crime, which would then have to be countered with a correspondingly modern criminal justice system. One might think of so-called *oracle attacks*, which criminals use to gain knowledge about the predictions of the corresponding predictive or big data policing software in order to adapt their criminal behavior accordingly (e.g., by breaking in exactly where the algorithm does not suspect a break-in). – Finally, the influence of AI in criminal law decision-making can also be trivialized. For example, by arguing that sentencing is too complex for machines to handle; or that in *State v. Loomis*, the Supreme Court ruled that algorithmic risk assessments may only be used to support and prepare independent judicial decisions, but may in no way bind or prejudge the latter.[72]

As engaging and comforting as criminal sociological trivializations of AI may sound, and as important as it is to keep in mind *de lega lata et ferenda* adaptation and displacement movements triggered by AI, criminal law theory must nevertheless deal more fundamentally with the promises of AI. Otherwise, it would involuntarily become their stirrup holder and miss the decisive "initial" phases of the upcoming developments. The more the weaknesses of AI are emphasized by criminal law theory, the more incentives are created for such weaknesses to be closed by technological progress. The trivialization of AI thus led, in substance, to a development spiral that creates social facts without being aware of their normative foundations.

For example: by correcting faulty code, *exploits* can be made increasingly impossible, *hacks* can be made more difficult by increasingly better firewalls, or *oracle attacks* can be anticipated (by predicting the abusive prediction of the regular predictions, with the consequence that patrols are carried out where they should not be). The importance of keeping such developments in mind from the outset in terms of criminal law theory is again demonstrated by *State v. Loomis.* For even if the decision-making process in criminal law is "only" to be prepared and supported by algorithms, this already generates – depending on

---

[72] On this the referenced decision in Fn. 61 above and in addition see Beriain, Iñigo De Miguel, «Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling», Law, Probability and Risk, 1/17 (2018) p. 47; Katyal, Sonia K., «Private Accountability in the Age of Artificial Intelligence», UCLA Law Review, 54/66 (2019) p. 86.

the layout – anchor effects[73] and a "compliance" pressure that should not be underestimated. If judges were allowed to overrule the algorithmic social prognosis with their own prognosis, this would lead to an actual prevalence of algorithmic methods (keyword: saving of working time; fear of negative reactions if one's own prognosis turns out to be wrong and e.g. the person released contrary to the algorithmic "advice" immediately recidivates), so that sentencers are enabled to shift responsibility ("blame shifting") to algorithms (according to the motto: "It's not me, but the machine that is responsible!"). These points must be well considered, which cannot be achieved by trivializing AI.

2. Undermining through Information Technology?

It is therefore more important to critically examine the promises of AI from the inside out and question their durability in terms of information technology, than to trivialize them from a perspective of criminal law theory. Admittedly, the challenges of AI cannot be met conclusively in this way, but only temporarily, if at all, in terms of criminal law theory. Here we encounter AI as the many-headed Hydra; as soon as one head is cut off, others grow back.

Nevertheless, it is necessary to express strong doubts about the promise of objectivity and neutrality of today's AI systems. Unrestrictedly neutral algorithms are difficult to imagine. Moreover, data-processing forecasts have to struggle with the so-called "*bias in, bias*" problem.[74]

a) Algorithmic Normativity

Data-processing algorithms work with self-learned or human determined criteria, e.g. when the age of an offender is given particular importance in calculating his risk of recidivism.[75] These algorithms become problematic, when their criteria are subject to conscious or unconscious normative targets – especially in the case of proprietary,

---

[73] On this and with regard to sentencing in general see Streng, Franz et al., *Strafgesetzbuch*, Baden-Baden: Nomos, 5. Ed., 2017, § 46 m.n. 3; Traut, Marcus; Nickolaus, Christoph, «Der Ankereffekt: Schattenseiten im Strafprozess», Strafverteidiger Forum, 12 (2015) p. 485.

[74] Following Mayson, Sandra G., «Bias in, Bias out», Yale Law Journal, 8/128 (2019) p. 2218.

[75] On this see Berk, Richard, «Algorithmic criminology», Security Informatics, 5/2 (2013) p. 4.

i.e. non-verifiable systems. As *Berk/Hyatt*, for instance, explain with regard to recidivism risk predictions – although linguistically somewhat dressed up, but in substance with thankful openness:

> "Many criminal justice stakeholders will treat false negatives as more costly than false positives. When this policy preferences applies, the standard of statistical proof necessarily will be lower for forecasts of homicide. The intent is to not release an individual who will commit a homicide and, in trade, to accept a larger number of false positives."[76]

This means that such a normatively oriented algorithm "approvingly accepts" (or more precisely: those behind the algorithm and those using it must approve of) leaving persons who are not dangerous (so-called false positives) in custody in order to prevent persons who are dangerous (so-called false negatives) from being erroneously classified as not dangerous and subsequently released from custody. It is obvious that such programming of criminal justice decisions – which is unfortunately not so far removed from German criminal law either, as the debates about preventive detention, for example, show[77] – is by no means neutral and objective, but rather highly politically and normatively charged.

In the literature, however, this is immediately turned into a positive aspect with critical intent, namely by linking it to the promise that AI forces the disclosure of normatively open and therefore politically fixable objectives.[78] Transparency thus becomes not only a basic requirement, but a normative good of the use of AI in the criminal justice system.

An example of this is *Chiao*'s recent (theoretical) discussion of algorithmic penalty assessment.[79] According to *Chiao*, an algorithm

---

[76] Hyatt, Jordan, «Machine Learning Forecasts of Risk to Inform Sentencing Decisions», Federal Sentencing Reporter, 4/27 (2015) p. 223.

[77] Cf. Boetticher, Axel et al., «Zum richtigen Umgang mit Prognoseinstrumenten durch psychiatrische und psychologische Sachverständige und Gerichte», Neue Zeitschrift für Strafrecht, 9/29 (2009) p. 478 (p. 479 with further references).

[78] As here for instance Beriain, Iñigo De Miguel, «Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling», Law, Probability and Risk, 1/17 (2018) p. 48 with further references.

[79] See Chiao, Vincent, «Predicting Proportionality: The Case for Algorithmic Sentencing», Criminal Justice Ethics, 3/37 (2018) p. 238 ff.

should evaluate the appropriateness of retrospective sentencing rather than prospective risk; that is, it should calculate not how dangerous a defendant is, but what sentence other judges in a given jurisdiction would impose in a comparable case. In this way, the judge who is actually called upon to assess the sentence is to be given a concrete guideline corridor. This proposal preempts traditional control mechanisms of sentencing (such as professional socialization processes, obligations of the courts to give reasons, and appellate reviews of appropriateness or arbitrariness). The idea is that sentencing is examined in individual cases for its systemic justice and correctness before it is pronounced in a legally binding manner. And since – as *Chiao* of course recognizes – the assessment of punishment is determined by the most diverse, partly antinomic goals and purposes, the algorithm to be applied would have to specify in a binding manner whether and which goals and purposes are to gain determining influence, and which weighting each is to have. With all this, *Chiao* opposes in a critical tradition the intransparency of the subjectively political side of the application of law. The transparency of an objective-politically designed algorithm is to take its place.

The algorithm thus becomes the *bouche de la loi*.[80] In other words, the belief of enlightenment in the ordering and pacifying power of rationality no longer connects with the human but with the machine-like *Subsumtionsautomat*,[81] which applies the norms of the legislator objectively, neutrally, and coherently.

b) Bias in, bias out

Moreover, the neutrality and objectivity of today's AI systems must be called into doubt by the *bias in, bias out*. This directly relates to the operation of AI-based (crime, recidivism risk, or *Chiao*'s sentencing) forecasts, which draw conclusions about the likelihood of future events by evaluating current data about past occurrences.[82] If, of course,

---

[80] Whether and how closely this figure is to be connected with Montesquieu may be left aside here.

[81] In general on the Figure of the *Subsumtionsautomat* see Ogorek, Regina, *Richterkönig oder Subsumtionsautomat? Zur Justiztheorie im 19. Jahrhundert*, Frankfurt a.M.: Vittorio Klostermann, 1986.

[82] An instructive case study from the USA is presented by Berk, Richard, «An Impact Assessment of Machine Learning Risk Forecast on Parole Board Decisions and Recividism», Journal of Experimental Criminology, 2/13 (2017) p. 193.

these past events are then prejudiced or reconstructed by prejudiced data sets, the prediction of the future reproduces the prejudices of the past in the present.[83] This is of immense relevance, especially in the U.S. discussion. The racial segregation that characterizes the U.S. criminal justice systems is represented by data sets that lead an algorithm to attribute a disproportionately high level of criminal energy to individual young African American males today that will be realized tomorrow, because this population group was disproportionately litigated (arrested, convicted, not released early from prison, etc.) through the criminal justice system yesterday. For which social (e.g. racist) reasons this took place "yesterday" remains algorithmically out of consideration from the beginning.[84]

Such a *bias in, bias out* does not have to be based on an error in the system, but can theoretically also have a system.[85] The promise of objectivity and neutrality would then contribute to algorithmically whitewashing a criminal justice system that has been tarnished (e.g., through racism), mystifying it as free of domination, and thus legitimizing it in social perception. All this can and should be criticized by a critical theory of criminal law that takes AI into account.

A way out of the information-technologically securitized dominance petrification of a *bias in, bias out* is promised by neutralized training data sets that reject normatively undesirable input factors (such as those that are directly or indirectly related to the skin color of the persons to be evaluated), as well as algorithms that normatively compensate for undesirable biases in the input data, by subtracting them.[86] Interestingly, this does not sustainably challenge the promise that AI

---

[83] See also, instructively Singelnstein, Tobias, «Predictive Policing: Algorithmenbasierte Straftatprognosen zur vorrausschauenden Kriminalintervention», Neue Zeitschrift für Strafrecht, 1/38 (2018) p. 4.

[84] In general and critical on this Hanna-Moffat, Kelly; Montford, Kelly Struthers, «Unpacking Sentencing Algorithms», in De Keijser, Jan W.; Roberts, Julian; Ryberg, Jesper, *Predictive Sentencing*, Oxford: Hart Publishing, 2018, p. 186 ff. With regard to the USA, Ferguson, Andrew G., «Illuminating Black Data Policing», Ohio State Journal Criminal Law, 2/15 (2018) p. 504 speaks of a "black data problem [since data is] racially encoded, colored by the history of real-world policing that disproportionality impacts communities of color.".

[85] This is brought into the discussion by Mayson, Sandra G., «Bias in, Bias out», Yale Law Journal, 8/128 (2019) p. 2218.

[86] Also critical on this and with further references: Mayson, Sandra G., «Bias in, Bias out», Yale Law Journal, 8/128 (2019) p. 2218.

could organize an objective, neutral as well as coherent criminal justice system more effectively and efficiently than human decision makers. On the contrary, it is reproduced at a higher level of order and issued as a development goal. Paradoxically, the criticism of the current use of AI in criminal justice outlined above stabilizes and legitimizes the future use of AI that is to be further developed.

The crucial doubts that the promises of AI are the proverbial "hot air" and that an AI-supported criminal justice system will virtually fail because of itself or because expectations are set too high, precisely because algorithms are programmed normatively and learn on the basis of biased data sets, are in the endless far-reaching than they might first appear. These doubts only concern the concrete implementation, but not the fundamental normative ideas of order of an algorithmic guarantee of an effective and efficient protection of legal interests as well as an objective, neutral and coherent application of criminal law. Rather, the demands for objective-political programming and for normative balancing of AI systems sanction and perpetuate these ideals. Ignoring, trivializing or undermining "AI and criminal law" is therefore of no help.

IV. REFLECTION ON FUNDAMENTAL NORMATIVE POSTULATES OF ORDER

In order to work through the transformative aspect of the connection between "AI and criminal law", a reflection on fundamental normative postulates of order is necessary – since ignoring, trivializing, and undermining are not sufficient; namely, the postulates of order of a desired criminal law as well as those of the desired society that is supposed to produce this criminal law. Since this raises large and major questions, only a few cursory considerations can be made here, but no definitive answers can be promised, so that in the following we must switch to the subjunctive at the crucial points.

1. Criminal Law as a Liberal Protection of Freedom or as a Welfare-State Security Law?

The real challenge for criminal law theory is that AI *prima facie* adopts – and optimizes – the central promises of criminal law.

If our "conventional human criminal law" (of any theoretical provenance) can only guarantee the protection of legal interests[87] normatively and counterfactually, because breaches of norms remain the order of the day, then AI wants to achieve the factual prevention or at least the substantial minimization of violations of legal interests in the long term. And can our "conventional human law" only provide a normative and counterfactual guarantee of objectivity and neutrality for legal decision-making, because personal idiosyncrasies and errors of the users of the law remain the "human" rule, with which the legal system has also come to terms (especially for pragmatic considerations),[88] AI wants to categorically prevent subjectivity, bias, and unequal treatment in the application of the law. If criminal laws would aim at being taken seriously by primary and secondary norm addressees nationwide,[89] in order to guarantee serious protection of the protected legal interests, and if the law were to take its claim to want to exhaustively guarantee objective, neutral, and coherent decision making seriously, then criminal law theory could not reject the promises of AI on principle. The factuality of (the promises of) AI would, in other words, resolve the end of the counterfactuality (of the promises) of criminal law – and thus the end in terms of the crowning conclusion of criminal law as we know it.

---

[87] On this paradigm see Hassemer, Winfried, *Theorie und Soziologie des Verbrechens*, Frankfurt a.M.: Athenäum, 1973, p. 27 ff.

[88] By way of example, it should be remembered that the legal protection guarantee in Article 19 section 4 of the German Basic Law only guarantees legal protection by, but not against, the judge, in accordance with the (admittedly controversial) so-called Dürig dogma.

[89] This qualification challenges us to deal more openly with the factual selectivity of criminal law – not only at the supra-national but also at the intra-national level – to think of the protection of legal interests under criminal law not only as fragmentary but as conceptually selective. After all, the degree of effectiveness that AI promises may quickly turn out to be too expensive for society in real life. For instance, in a Chinese newspaper the AI-based anti-corruption system there, which significantly bears the zero trust maxim in its name, was critically questioned as follows: "Is China's corruption-busting AI system 'Zero Trust' being turned off for being too efficient?" The background is the concern that the Chinese public administration would not be able to cope with a comprehensive prosecution of all identified corruption offenses. Access: https://www.scmp.com/news/china/science/article/2184857/chinas-corruption-busting-ai-system-zero-trust-being-turned-being [Date of consultation 14.12.2021].

A fundamental criticism of the entry of AI into the criminal justice system would therefore have to take a critical look not only at its "foreign" objectives, but also at its "own" objectives, which are supposedly only taken over by AI. To defend and justify would be, in other words, nothing less than the "only" normative and counterfactual aspects of criminal justice. These are: the real possibility of criminal violations of legal rights as well as the real possibility of biased, prejudiced, idiosyncratic human law practitioners.[90] At this point, we encounter the difference between a liberal foundation of criminal law, which dedicates the latter to the protection of freedom, and welfare-state conceptions, which (can) use criminal law to protect security.[91]

In order to be able to counter the promises of AI in principle, the dogma of the "protection of legal rights through criminal law" would have to be supplemented from a *liberal perspective* – and since we cannot give definitive answers here, the subjunctive can and must be used in the following. This could be achieved – (partly very!) loosely according to *Haffke*,[92] *Tiedemann*[93] and *Prittwitz*[94] –, by dedicating criminal law fundamentally to the protection of liberty, to which other (possibly legitimate, but then not in this sense criminal-legal) instruments for the prevention of violations of legal interests are to be seen in contrast to. What is meant by this is the direct protection of a use of freedom (such as the consumption of alcohol free from state supervision and control), which indirectly also enables the abuse of freedom, i.e. the freedom to commit criminal acts (such as negligent or intentional driving under the influence of alcohol).

---

[90] E.g. by placing law in the service of discursive, justification- and critique-driven orientation toward objectivity, neutrality and coherence.

[91] This ideal-typical juxtaposition is also found in Günther, Klaus, «Bedrohte individuelle Freiheiten im aufgeklärten Strafrecht – Welche Freiheiten?», Kritische Justiz, 4/49 (2016) p. 520.

[92] Haffke, Bernhard, «Die Legitimation des staatlichen Strafrechts zwischen Effizienz, Freiheitsverbürgung und Symbolik», in Schünemann, Bernd et al., *Festschrift für Roxin zum 70. Geburtstag*, Berlin: De Gruyter, 2001, p. 965.

[93] Concisely summarized in Tiedemann, Klaus, *Wirtschaftsstrafrecht: Einführung und Allgemeiner Teil*, 5. Ed., Munich: Vahlen, 2017, m.n. 228.

[94] Prittwitz, Cornelius, «Strafrecht als propria ratio», in Heinrich, Manfred et al., *Strafrecht als Scientia Universalis: Festschrift für Claus Roxin zum 80. Geburtstag*, Berlin/New York: De Gruyter, 2011, p. 23 ff.

Only in this way could the non-commitment of a crime continue to be evaluated as a free decision for and its commission as a free decision against the law from an *individual perspective*. In a (hypothetical) world in which AI makes crimes *eo ipso* or *de facto* impossible, there can no longer be any question of this freedom, the famous ability to act differently, even if only as a fiction necessary for a liberal community.

Moreover, from a *social point of view*, even up to now (i.e. without AI) ways and means were conceivable with which violations of legal rights could be made *eo ipso* impossible or at least *de facto* drastically (and draconically) minimized.[95] The potentials of AI culminate in this sense what was also conceivable up to this point in the context of a rigid welfare-state, namely administrative or "technical prevention" (*Hassemer*[96]), in particular a kind of "technological paternalism" (*Hilgendorf*).

If the counterfactual of criminal law is to be preferred to the factual of more effective instruments of crime prevention, a liberal theory of criminal law would have to deal more openly with balancing liberties. The loss of freedom of the many, who have to submit to rigid non-criminal measures, although they can also be reached by normative (criminal) measures, would then have to weigh more heavily than the possible loss of freedom of the few, whose legal interests are violated by those who were not normatively (criminally) addressable. In terms of penal constitutional law: penal and prohibition norms would have to be theorized as *prima ratio* of the protection of liberty, because and if a more rigid (e.g., administratively supervisorial, regulative, or technically intervening) protection of legal rights would, on balance, not be necessary or appropriate, since it would excessively burden individuals or the general public.

---

[95] In criminal economic law, for example, through rigid administrative supervision and regulation of economic operators. Or, in criminal traffic law, through technical intervention against drivers. One might think here of mandatory alcohol tests before driving, linked to an immobilizer. This is not (!) science fiction either, but is being considered concretely (!) at EU level. On this see https://www.europarl.europa.eu/news/en/press-room/20190410IPR37528/parliament-approves-eu-rules-requiring-life-saving-technologies-in-vehicles [Date of consultation 14.12. 2021] with further references.

[96] Hassemer, Winfried, «Aktuelle Perspektiven der Kriminalpolitik», Strafverteidiger, 6 (1994), p. 333 ff. (p. 336).

Illustrative of this is the reciprocal freedom balance that is opened up, for example, in the case of high-risk technology (think, for example, of commercial nuclear power). Here, the possible loss of freedom of the many (which would have to be feared, for example, in the case of a nuclear total meltdown) outweighs the loss of freedom of the few, who are subjected to rigid non-criminal measures (such as close-meshed state control of the operators of nuclear power plants). The painful price of this reconstruction is that the victims of crime have to accept their real loss of freedom in order to secure for the others their virtual other preservation of freedom. To put it bluntly: the parents of a child who was run over because a driver negligently relied on the approved "autopilot" of his car must be told openly that the ban on such "autopilots" was socially too "expensive", e.g. too anti-innovation, and also not in the sense of the many who use these "autopilots" in a traffic-friendly way to realize their freedom.

A liberal criminal law to be legitimized in this way could be positioned against the promise of an effective as well as efficient protection of legal rights by AI. Of course, this would require a great deal of courage in terms of criminal policy and penal theory. After all, the public and private third-party and self-monitoring that is necessary to enable predictive and big data policing would have to be classified as excessively invasive. And the virtual gain of freedom of the many, who are not exposed to surveillance, would have to be preferred to the real loss of freedom of the few, who become victims of criminal acts, which could (probably) have been prevented by the use of AI.

The necessary trade-offs and balancing of freedoms will have to be made on a sector-specific basis in the future as well. The penal law as a liberal law for the protection of liberty would therefore not have to be applied everywhere and in general. The liberal theory of criminal law, however, is required to weigh freedom and the protection of rights in an open process.

2. Which Criminal Law for which Society?[97]

The pressure to show one's colors also arises as soon as one takes a look at AI's underlying normative postulates of order, which are easily

[97] This formulation of my question originates from discussions in the circle of colleagues in Frankfurt and I ultimately owe it to *Klaus Günther*.

obscured by many superficial justificatory narratives (more effective and efficient; more objective, neutral, and coherent). In this respect, criminal legal theory must open up to social or political theory and take a stand on fundamental issues (in particular, the social status of trust and openness to the future).

As has been shown (above II.), the use of AI in criminal justice is and would be an expression of a fundamental loss of interpersonal trust. Any other person (including the user of the law) is no longer to be trusted; but rather, to be managed as a potential danger and as a risk to be monitored and whose future behavior is to be algorithmically anticipated. In its own logic, this amounts to a generalization of automated suspicion – or, in terms of criminal procedure, of automated and general initial suspicion.[98]

In order to oppose the promises of AI in principle, the presumption of innocence would have to be held higher in social-criminal theory, namely more comprehensively as a counterfactual interpersonal presumption of trust and thus, e.g., would have to, quite enlightenment-like – as *Hruschka* has clearly worked out –, be reconstructed as "everyone's dignity".[99] This presupposes the closing of ranks with a social philosophy that does not regard trust as a mere mechanism for reducing social complexity (*Luhmann*), but as functionally valuable in order to prevent a slide into a trustless surveillance society with an always potentially authoritarian and oppressive character.[100]

It is no less challenging to meet the ideal of an *end of history*.[101] The functioning of AI proclaims, as we have seen (above II.), in the matter a kind of end of history, simultaneously a closure of the future. After all, conclusions about the future are drawn from the past, which

---

[98] Brennan-Marquez, Kiel, «Big Data Policing and the Redistribution of Anxiety», Ohio State Journal of Criminal Law, 2/15 (2018) p. 488 therefore aptly diagnoses that under the impression of predictive and big data policing, the constitutional figure of initial suspicion, which legitimizes state encroachments on fundamental rights, is being worn away (beyond recognition).

[99] Hruschka, Joachim, «Die Unschuldsvermutung in der Rechtsphilosophie der Aufklärung», Zeitschrift für die gesamte Strafrechtswissenschaft, 2/112 (2000) p. 285.

[100] Whether the concept of trust as developed by Parsons, Talcott, *Politics and Social Structure*, New York: Free Press, 1969, which represents a political concept of order in response to the Hobbesian situation, can be used for this purpose does not need to be further explained here.

[101] Downright classic: Fukuyama, Francis, *The End of History and the Last Man*, New York: Free Press, 1992.

– for better or worse – leads to a petrification of yesterday in today and inhibits the dynamic development of tomorrow. The promise is that an abnormal future can actually be prevented by means of AI.

At first glance, this is also the aim of traditional criminal law theory. The individual is relieved of the concern about future violations of legal rights because and by guaranteeing the future existence of these legal rights in the present by means of criminal law. The difference is once again that AI promises a factual and criminal law a counterfactual or normative *end of history*. This has consequential effects. In an ideal (utopian or dystopian) AI world, dissidence and resistance against the *status quo* reproduced as *status quo ante* are not only futile (because and if they are made *eo ipso* or *de facto* impossible), but ideally also inconceivable (especially because and if the individual makes himself the executor of his own subjection to a panoptic form of rule).

Whoever wanted to reject this closure of the future and advocate its opening could find the intrinsic and added value of the counterfactual of criminal law in the *de facto* admission of dissidence and resistance. Deviant behavior should then no longer be labeled solely as a violation of legal rights that needs to be prevented, but should at least also be recognized as potential (objective or subjective) criticism of the *status quo* (one need only think of the criminal prohibition of homosexual intercourse, which was brought down not least by continuous acts of resistance, namely by acts of norm-breaking). This meant that in the discussion of the violation of norms, the criminal reaction would have to be justified again and again with good reasons and that it could not, for example, be presented as "natural". Criminal law would thus have to be conceived as an evolutionary and discursive practice of (human) justification and critique, and the temporal contingency of criminal law would have to be accepted in order to contrast the open future of criminal law with the closed future of AI.

OUTLOOK

This contribution advocates for viewing AI as a social construct whose practice is capable of transforming our fundamental social and (criminal) legal notions of order. In light of concrete criminal law applications, AI as a transformative technology should no longer be ignored

or trivialized in terms of criminal law theory. And since even internal information-technological doubts about the performance of AI only promote further spirals of development, it is necessary to critically question the normative postulates of order that are programmed into AI. In particular, it must be questioned what the price is of AI's promise, that it is able to guarantee a more effective and efficient protection of legal rights and a more neutral, objective and coherent application of criminal law (keyword: victory of the welfare-state security and *zero trust* paradigm in a Big Data-based surveillance society). Conversely, a non-ideal theory of criminal law would – as can only be stated here in conclusion – oversimplify its opposition to the entry of AI into the criminal justice system, by countering it with (normative) ideals (keyword: criminal law as freedom protection law; law as discursive practice of justification and criticism).[102] For this would presuppose a particularly sophisticated criminal law that is (and should not be "merely") free of all authoritarian and oppressive borrowings.[103] Anyone who wants to present this as a desirable goal, but one that is not very realistic in terms of critical intent[104], and who must assume that the digitization of society will in fact continue to advance, will ultimately have to think about how criminal law, which in principle must be liberal, democratic and based on the rule of law, can be supplemented by AI systems in such a way that AI does not corrupt criminal law in an authoritarian manner on the one hand, and on the other hand can free it from illiberal tendencies that are contrary to the rule of law and authoritarian.[105]

---

[102] To hint at the famous methodological debate between ideal and non-ideal theory formation for criminal law theory, as set off by *Rawls*.

[103] As such in general, without a specific reference to AI, Naucke, Wolfgang, *Negatives Strafrecht*, Berlin: Lit Verlag, 2015, p. 114 ff.

[104] Ideal theory formation is exposed to the criticism that it stabilizes the status quo. A non-ideal theory formation does not let itself be driven by current developments, but takes note of them and wants to understand them in order to make them the point of reference for critical reflections.

[105] E.g. by using AI for educational purposes in order to show judges their prejudices, etc. On this impressively Sommer, Ulrich, «Psychologie der richterlichen Entscheidungsfindung», Zeitschrift für Rechtspolitik, 2/50 (2017) p. 60, who considers scientific research into judicial prejudices to be necessary because it is still "taboo" in this country. In terms of legal methodology and sentencing methodology, the work of Hruschka, Joachim, «Rechtsanwendung als methodisches Problem», Archiv für Rechts- und Sozialphilosophie, 4/50 (1964), p. 485 (especially p. 498) continues to be extremely worth reading.

In this (non-ideal and critical) sense, AI would then not have to be designed as the end of criminal law (neither in the sense of the dying death of a liberal criminal law of liberty nor in the sense of the crowning conclusion of a welfare-state criminal law of security), but as a building block of a criminal law of the (near) future to be designed today, which at the same time shows itself to be open to technology[106] *and* humane and thus, continues to be reasonable in a modern sense. In all of this, it is important to heed *Joachim Hruschka's* great legacy – which is not always easy to implement, especially in view of AI – that criminal law must always remain *intellectually honest*, i.e. it must use rational argumentation and pedantically avoid incantations and magic formulas.[107]

REFERENCES

Balkin, Jack B., «The Path of Robotics Law», California Law Review Circuit, 6 (2015) p. 45-60.

Beriain, Iñigo De Miguel, «Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling», Law, Probability and Risk, 1/17 (2018) p. 45-53.

Berk, Richard; Hyatt, Jordan, «Machine Learning Forecasts of Risk to Inform Sentencing Decisions», Federal Sentencing Reporter, 4/27 (2015) p. 222-228.

---

[106] Attempts to close criminal law to technological innovations in law can seem artificial at best. For example, French criminal law now includes the following provision: "Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées." Access, and with reference to where this is to be implemented: https://www.legifrance.gouv.fr/eli/ loi/2019/3/23/JUST1806695L/jo/texte [Date of consultation 14.12. 2021]. – This raises the concern of making judicial decision-makers algorithmically "transparent". If this should be associated with the concern that it can be made clear algorithmically that these decision-makers do not apply the law objectively and neutrally, but rather make decisions colored by subjectivity, the question naturally arises as to whether the protection of a myth or mere ideal by criminal law (i.e., the objectivity and neutrality of the application of the law) or of an unquestionable trust in judicial decision-makers represents a legitimate and meaningful legal interest.

[107] As explicitly stated by Hruschka, Joachim, *Strafrecht nach logisch-analytischer Methode*, Berlin/New York: De Gruyter, 1988, p. XVIII.

Berk, Richard, «An Impact Assessment of Machine Learning Risk Forecast on Parole Board Decisions and Recividism», Journal of Experimental Criminology, 2/13 (2017) p. 193-216.

Berk, Richard, «Algorithmic criminology», Security Informatics, 5/2 (2013) p. 1-14.

Boetticher, Axel et al., «Zum richtigen Umgang mit Prognoseinstrumenten durch psychiatrische und psychologische Sachverständige und Gerichte», Neue Zeitschrift für Strafrecht, 9/29 (2009) p. 478-481.

Botnick, Claire, «Evidence-based Practice and Sentencing in State Courts: A Critique of the Missouri System», Washington University Journal of Law & Policy, 1/49 (2015) p. 159-180.

Brantingham, Jeffrey, «The Logic of Data Bias and Its Impact on Place-Based Predictive Policing», Ohio State Journal of Criminal Law, 2/15 (2018) p. 473-486.

Brennan-Marquez, Kiel, «Big Data Policing and the Redistribution of Anxiety», Ohio State Journal of Criminal Law, 2/15 (2018) p. 487-493.

Byrd, Sharon; Hruschka, Joachim, «Kant zu Strafrecht und Strafe im Rechtsstaat», Juristen Zeitung, 20/62 (2007) p. 957-1008.

CEPEJ, «European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment» v. 3 – 4.12.2018, in particular p. 64 ff. [Date of consultation 14.12. 2021], Access: https://rm.coe.int/ ethical-charter-en-for-publication-4-december--2018/16808f699c.

Chiao, Vincent, «Predicting Proportionality: The Case for Algorithmic Sentencing», Criminal Justice Ethics, 3/37 (2018) p. 238-61.

Deskus, Cassie, «Fifth Amendment Limitations on Criminal Algorithmic Decision-Making», NYU Journal of Legislation and Public Policy, 1/21 (2018) p. 237-286.

Durkheim, Emile, *Die Regeln der soziologischen Methode*, Frankfurt a.M.: Suhrkamp, 1984.

Ertel, Wolfgang, *Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung*, Wiesbaden: Springer VS, 2016.

European Committee on Crime Problems (CDPC), «Artificial Intelligence and its Impact on CDPC Work: The case of automated driving», CDPC (2018) 14 v. 14. 09. 2018, p. 6 f. [Date of consultation 14.12. 2021], Access: https://rm.coe.int/cdpc-2018

-14-artificial-intelligence-and-criminal-law-project-2018-202/16808d6d09.

Fateh-Moghadam, Bijan, «Selbstbestimmung im biotechnischen Zeitalter», Basler Juristische Mitteilungen, 4 (2018) p. 205-232.

Ferguson, Andrew G., «Illuminating Black Data Policing», Ohio State Journal Criminal Law, 2/15 (2018) p. 503-525.

Ferguson, Andrew G., *The Rise of Big Data Policing. Surveillance, Race, and the Future of Law Enforcement*, New York: New York University Press, 2017.

Forst, Rainer; Günther, Klaus, «Die Herausbildung normativer Ordnungen: zur Idee eines interdisziplinären Forschungsprogramms», in Forst, Rainer/Günther, Klaus, *Die Herausbildung normativer Ordnungen*, Frankfurt a.M.: Campus, 2011, p. 11-30.

Fries, Martin, «Automatische Rechtspflege», Rechtswissenschaft, 4/9 (2018) p. 414-430.

Fry, Hannah, *Hello World: Was Algorithmen können und wie sie unser Leben verändern*, München: C.H. Beck, 2018.

Fukuyama, Francis, *The End of History and the Last Man*, New York: Free Press, 1992.

Gaede, Karsten, *Recht und Strafen für Roboter?*, Baden-Baden: Nomos, 2019.

Günther, Klaus, «Bedrohte individuelle Freiheiten im aufgeklärten Strafrecht – Welche Freiheiten?», Kritische Justiz, 4/49 (2016) p. 520-534.

Haffke, Bernhard, «Die Legitimation des staatlichen Strafrechts zwischen Effizienz, Freiheitsverbürgung und Symbolik», in Schünemann, Bernd et al., *Festschrift für Roxin zum 70. Geburtstag*, Berlin: De Gruyter, 2001, p. 955-977.

Han, Byung-Chul, *Psychopolitik: Neoliberalismus und die neuen Machttechniken*, Frankfurt a.M.: Fischer, 2015.

Hanna-Moffat, Kelly; Montford, Kelly Struthers, «Unpacking Sentencing Algorithms», in De Keijser, Jan W.; Roberts, Julian; Ryberg, Jesper, *Predictive Sentencing*, Oxford: Hart Publishing, 2018, p. 175-196.

Hassemer, Winfried, «Aktuelle Perspektiven der Kriminalpolitik», Strafverteidiger, 6 (1994) p. 333-336.

Hassemer, Winfried, *Theorie und Soziologie des Verbrechens*, Frankfurt a.M.: Athenäum, 1973.

Heckmann, Jörn, «DAO- Hack: smart contracts auf dem rechtlichen Prüfstand», Computer und Recht, 9 (2016) R99-R100.

Henderson, Stephen E., «A Few Criminal Justice Big Data Rules», Ohio State Journal of Criminal Law, 2/15 (2018) p. 527-541.

Herberger, Maximilian, «„Künstliche Intelligenz" und Recht», Neue Juristische Wochenschrift, 39/71 (2018) p. 2825-2829.

Hildebrandt, Mireille, *Smart Technologies and the End(s) of Law*, Cheltenham: Edward Elgar Publishing, 2015.

Hilgendorf, Eric, «Autonome Systeme, künstliche Intelligenz und Roboter», in Barton, Stephan et al., *Festschrift für Thomas Fischer*, Munich: C.H. Beck, 2018, p. 99-113.

Hruschka, Joachim, «Der Standard-Fall der aberratio ictus und verwandte Fallkonstellationen», Juristen Zeitung, 10/46 (1991) p. 488-492.

Hruschka, Joachim, «Die Herbeiführung eines Erfolges durch einen von zwei Akten bei eindeutigen und bei mehrdeutigen Tatsachenfeststellungen», Juristische Schulung (1982) p. 317-325.

Hruschka, Joachim, «Kann und sollte die Strafrechtswissenschaft systematisch sein?», Juristen Zeitung, 1/40 (1985) p. 1-10.

Hruschka, Joachim, *Strafrecht nach logisch-analytischer Methode*, Berlin/New York: De Gruyter, 1988.

Hruschka, Joachim, «Das Strafrecht neu durchdenken!», Goltdammer's Archiv (1981) p. 237-250.

Hruschka, Joachim, «Die Unschuldsvermutung in der Rechtsphilosophie der Aufklärung», Zeitschrift für die gesamte Strafrechtswissenschaft, 2/112 (2000) p. 285-300.

Hruschka, Joachim, «Rechtsanwendung als methodisches Problem», Archiv für Rechts- und Sozialphilosophie, 4/50 (1964), p. 485--501.

Hruschka, Joachim, «Utilitarismus in der Variante von Peter Singer», Juristen Zeitung, 6/56 (2001) p. 261-271.

Huq, Aziz Z., «Racial Equity in Algorithmic Criminal Justice», Duke Law Journal, 6/68 (2019) p. 1043-1134.

Irsigler, Ingo; Orth, Dominik, «Zwischen Menschwerdung und Weltherrschaft: Künstliche Intelligenz im Film», Aus Politik und Zeitgeschichte, 6-8/68 (2018) p. 39.

Katyal, Sonia K., «Private Accountability in the Age of Artificial Intelligence», UCLA Law Review, 54/66 (2019) p. 56-141.

Katz, Yarden, «Manufacturing an Artificial Intelligence Revolution», SSRN (2017) p. 1, [Date of consultation 14.12. 2021], Access: https://ssrn.com/abstract=3078224.

Lewis, Nell, «Should AI be used to catch shoplifters?», *CNN Business* (2019) [Date of consultation 14.12.2021], Access: https://edition.cnn.com/2019/04/18/business/ai-vaak-shoplifting/index.html.

Liptak, Adam, «Sent to Prison by a Software Program's secret Algorithms», *New York Times*, May 1, 2017.

Liu, Han-Wei; Lin, Ching-Fu; Chen, Yu-Jie, «Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability», International Journal of Law and Information Technology, 2/27 (2019) p. 122-141.

Luhmann, Niklas, *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität*, 5. Ed., Munich: UVK, 2014.

Martini, Mario; Nink, David, «Wenn Maschinen entscheiden … – vollautomatisierte Verwaltungsverfahren und der Persönlichkeitsschutz», Neue Zeitschrift für Verwaltungsrecht – Extra, 10/36 (2017) p. 1-14.

Mau, Steffen, *Das metrische Wir: Über die Quantifizierung des Sozialen*, Berlin: Suhrkamp, 2017.

Mayson, Sandra G., «Bias in, Bias out», Yale Law Journal, 8/128 (2019) p. 2122-2473.

McEwan, Ian, *Machines like Me*, London: Penguin, 2019.

Meng, Yu; Guodong, Du, «Why Are Chinese Courts Turning to AI?», The Diplomat (2019) [Date of consultation 14.12. 2021], Access: https://thediplomat.com/2019/01/why-are-chinese-courts-turning-to-ai/.

Meyer, Stephan, «Künstliche Intelligenz und die Rolle des Rechts für Innovation», Zeitschrift für Rechtspolitik, 8/51 (2018) p. 233-238.

Nakamoto, Satoshi, «Bitcoin: A Peer-to-Peer Electronic Cash System» (2009) p. 1-8 [Date of consultation 14. 12. 2021], Access: https://bitcoin.org/bitcoin.pdf.

Naucke, Wolfgang, *Negatives Strafrecht*, Berlin: Lit Verlag, 2015.

Nida-Rümelin, Julia; Weidenfeld, Nathalie, *Digitaler Humanismus*, Munich: Piper, 2018.

Ogorek, Regina, *Richterkönig oder Subsumtionsautomat? Zur Justiztheorie im 19. Jahrhundert*, Frankfurt a.M.: Vittorio Klostermann, 1986.

Ostermeier, Lars, «Der Staat in der prognostischen Sicherheitsgesellschaft», in Puschke, Jens; Singelnstein, Tobias, *Der Staat und die Sicherheitsgesellschaft*, Wiesbaden: Springer VS, 2017, p. 101-121.

Palka, Silvia; Wittpah, Volker, «Vertrauen und Transparenz – Blockchain Technologie als digitaler Vertrauenskatalysator», Working Paper of the Institute for Innovation and Technology, 39 (2018) p. 1-15.

Parsons, Talcott, *Politics and Social Structure*, New York: Free Press, 1969.

Prittwitz, Cornelius, «Strafrecht als propria ratio», Heinrich, Manfred et al., *Strafrecht als Scientia Universalis: Festschrift für Claus Roxin zum 80. Geburtstag*, Berlin/New York: De Gruyter, 2011, p. 23-39.

Rademacher, Timo, «Predictive Policing im deutschen Polizeirecht», Archiv des öffentlichen Rechts, 3/142 (2017) p. 366-416.

Ramge, Thomas, *Mensch und Maschine: Wie künstliche Intelligenz und Roboter unser Leben verändern*, Stuttgart, Reclam, 2018.

Schemmel, Alexander; Dietzen, Alexandra, «"Effective Corporate Governance" by Legal Tech & Digital Compliance», in Breidenbach, Stephan/Glatz, Florian, *Rechtshandbuch Legal Tech*, Munich: Beck, 2018, p. 137-152.

Schwintowski, Hans-Peter, «Wird Recht durch Robotik und künstliche Intelligenz überflüssig?», Neue Juristische Online-Zeitschrift, 42 (2018) p. 1601-1609.

Searle, John R., «Minds, Brains, and Programs», Behavioral and Brain Sciences, 3/3 (1980) p. 417-457.

Singelnstein, Tobias, «Predictive Policing: Algorithmenbasierte Straftatprognosen zur vorrausschauenden Kriminalintervention», Neue Zeitschrift für Strafrecht, 1/38 (2018) p. 1-9.

Sommer, Ulrich, «Psychologie der richterlichen Entscheidungsfindung», Zeitschrift für Rechtspolitik, 2/50 (2017) p. 60-62.

Straube, Till; Belina, Bernd, «Policing the Smart City: Eine Taxonomie polizeilicher Prognoseprogramme», in Bauriedl, Sybille; Strüver, Anke, *Smart City – Kritische Perspektiven auf die Digitalisierung in Städten*, Bielefeld: transcript Verlag, 2018, p. 223-236.

Streng, Franz et al., *Strafgesetzbuch*, Baden-Baden: Nomos, 5. Ed., 2017.

Tiedemann, Klaus, *Wirtschaftsstrafrecht: Einführung und Allgemeiner Teil*, 5. Ed., Munich: Vahlen, 2017.

Traut, Marcus; Nickolaus, Christoph, «Der Ankereffekt: Schattenseiten im Strafprozess», Strafverteidiger Forum, 12 (2015) p. 485-492.

Wagner, Gerald, «Vertrauen in Technik», Zeitschrift für Soziologie Zeitschrift für Soziologie, 2/23 (1994) p. 145-157.

Weizenbaum, Joseph, *Die Macht der Computer und die Ohnmacht der Vernunft*, Frankfurt a.M.: Suhrkamp, 1977.

Xanke, Lisa; Bärenz, Elisabeth, «Künstliche Intelligenz in Literatur und Film – Fiktion oder Realität?», Journal of New Frontiers in Spatial Concepts, 4 (2012) p. 36-43.

Zuboff, Shoshana, *Das Zeitalter des Überwachungskapitalismus*, Frankfurt a.M./New York: Campus, 2018.

ISBN 978-989-9075-19-1

FCT

Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR