Transparency for contesting automated decisions

Bayamlioğlu, Emre

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Transparency for Contesting Automated Decisions

Impediments and Affordances under EU Law

Emre Bayamlıoğlu

Transparency for Contesting Automated Decisions

Impediments and Affordances under EU Law

Emre Bayamlıoğlu

Transparency for Contesting Automated Decisions

Impediments and Affordances under EU Law

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, en Vrije Universiteit Brussel op gezag van de rector magnificus, prof. dr. J. Danckaert in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Aula van Tilburg University op dinsdag 13 juni 2023 om 16.30 uur

door

Ibrahim Emre Bayamlıoğlu,

geboren te Ankara, Turkije.

Promotores:	Prof. Dr. R. Leenes	(Tilburg University)
	Prof. Mr. Dr. M. Hildebrandt	(Vrije Universiteit Brussel)

Promotiecommissie:	Prof. Mr. Dr. S. Zouridis	(Tilburg University)
	Dr. O. Lynskey	(London School of Economics)
	Prof. Dr. G. González Fuster	(Vrije Universiteit Brussel)
	Prof. Dr. M. Birnhack	(Tel Aviv University)

Contents

Summary

The thesis, answering the research question: *How could transparency of automated decision-making be constructed as a prerequisite of contestation and what could be the main affordances and impediments under EU law?* consists of four peer-reviewed articles.

The research does not provide a full account of the entire regulatory landscape of transparency but rather focuses on essential requirements of the concept in the context of automated decision-making (ADM) — analysing the relevant provisions of the *GDPR* and *intellectual property* laws as the most prominent body of rules relevant to the legal scrutiny and contestation of ADM systems.

Since the thesis is focused on a specific type of transparency, that is, an actionable one for the purposes of contestation, the first paper "*The 'Rule of Law' implications of data-driven decision-making: a techno-regulatory perspective"* [1] (co-authored with R. Leenes) sets the ground by conceptualising data-driven ADM as the new horizon of techno-regulation. The paper puts forward the perspective that transparency in automated decisions is not about reading of computer code but rather relates to the question how these systems make up the normative landscape that we are subjected to. As an extension of Lessig's "Code as Law", the thesis rests on the premise that by *sorting, classifying* and *predicting* these technologies impose or facilitate certain norms, values or criteria. As such, the paper defines *automated decision-making* (ADM) as a regulatory technology and identifies three impairments (normative, causal, moral) which undermine the principal of rule of law. This theoretical stance allows for a conceptualisation of ADM and the surrounding transparency debate as a procedural, or we may say, as a *due process* problem.

Next, the thesis develops a transparency model, laying out the required forms and degrees of transparency necessary to contest automated decisions. The *second paper*[2] initially explains how i) technical complexities; ii) epistemological flaws (spurious correlation or weak causation); and iii) biased processes inherent in machine learning (ML) create obstacles in terms of interpreting automated decisions. The analysis illustrates that a conception of transparency aiming to see the entire system 'at work' is an ever-expanding territory, that is, as you open black-boxes, you may just find more black-boxes. Instead, *the* outcome of ML-based systems may eventually be attributed to the values and assumptions that underlie the response of the system to a given input. Accordingly, the transparency (contestation) model developed in the second paper is a reconstruction of ADM as a 'rule-based' process where certain input lead to certain results—akin to the decisions in a legal system based on *facts*, *norms* and the ensuing *consequences*.

Having identified the essential transparency requirements for effective contestation, the remainder of the thesis focuses on the relevant legal frameworks. implementation of the transparency model—exploring to what extent the relevant provisions in the GDPR could be interpreted in the direction of "contestability". For this purpose, the *third paper*[3] provides a systematic and teleological interpretation of Article 22 of the GDPR on automated decisions—focussing on the question how the rights to obtain human intervention, express one's views and contest the decision could practically be implemented. By defining Art 22 as a general provision of *due process* and the right to contest as the core remedy provided by the GDPR against ADM, the paper transcends the current debates about the existence and the scope of a so-called "right to an explanation".

---

[1] "The "Rule of Law" Implications of data-driven decision-making: A techno-regulatory perspective", LIT (2018) 10:2, 295-313, 10.1080/17579961.2018.1527475 (Co-author R. Leenes)

[2] "Contesting automated decisions: A view of transparency implications", EDPL (2018), 4(4), 433-446. https://doi.org/10.21552/edpl/2018/4/6

[3] "The right to contest automated decisions under the GDPR: Beyond the so-called right to explanation", *Regulation & Governance;* 2021; Vol. 15, Special Issue*: Algorithmic Regulation* eds. by Karen Yeung and Lena Ulbrich 10.1111/rego.12391

As the most important legal framework that could impede transparency efforts, the final part of the thesis on IP rights provides a macro-view of the potential areas of conflict between the transparency requirements and the relevant IP regimes—i.e., copyright, sui generis database right and trade secret protection. The *fourth paper*[4] initially clarifies that the implementation of transparency measures and mechanisms as defined in the previous parts of the thesis require the disclosure, reproduction or modification of certain informational elements of ML systems. Following this, the paper explores i) to what extent reliance on IP rights could excuse ADM from the obligation of making transparent and contestable decisions, e.g., under Article 22 of the GDPR and ii) what are the counter-arguments based on statutory exceptions and limitations restricting IP rights. The paper analyses the IP-eligible elements in ML-based systems in a dual structure as: data and datasets (*expressional elements*) on one side and algorithmic techniques and ML models (*utilitarian/ operational elements*) on the other.

[4] "Machine Learning and the relevance of IP rights: An account of transparency requirements for AI", *Forthcoming* EPLR Special Issue on AI*, 2023*

Abbreviations

ACM	Association for Computing Machinery
ADM	Automated decision-making
AI	Artificial Intelligence
AIA	Algorithmic Impact Assessment
API	Application programming interface
B2B	Business-to-business
B2G	Business-to-government
CbD	Contestation by DEsign
CJEU	The Court of Justice of the European Union
DA	Data Act
DGA	Data Governance Act
DPD	Data Protection Directive
DPIA	Data Protection Impact Assessment
DPbD	Data Protection by Design and by Default
DRM	Digital Rights Management
DSM	Digital Single Market (Directive)
EU	European Union
EDPS	European Data Protection Supervisor
EPC	European Patent Convention
EPO	European Patent Organisation
FS	Feature selection
FE	Feature extraction
ICT	Information and Communication Technology
IoT	Internet of Things
IP	Intellectual Property
ISO	International Organization for Standardization
IEEE-SA	Institute of Electrical and Electronics Engineers Stan- dards Association
GDPR	General Data Protection Regulation
LKBS	Legal knowledge based systems
ML	Machine Learning
NLP	Natural Language Processing
RbM	Rule-based Model
RFID	Radio-frequency identification
RQ	Research Question
STS	Science and Technology Studies
TDM	Text and data mining
TPM	Technological protection measure
TS	Trade Secret
TRIPs	The Agreement on Trade-Related Aspects of Intellectual Property Rights
UGC	User-generated Content
WIPO	The World Intellectual Property Organization
XAI	Explainable Artificial Intelligence

Chapter 1

Introduction: The Design of the Research Project

The only way to rectify our reasonings is to make them as tangible as those of the Mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: Let us calculate, without further ado, in order to see who is right.[*]

[*] Leibniz, *Selections*. Philip P. Wiener(ed.), New York: Charles Scribner's Sons, 1979, 51.

Introduction: The Design of the Research Project

1. Preliminaries and the background

As the industrial "revolution" was based on the modelling of machines for specific mechanical tasks, the new era of "computational turn" is characterized by its modelling of processes from manufacturing of goods to simulation of real-life scenarios— and even the "maddening randomness of humans"[1]— extending the physical assembly line of *Henry Ford* to a virtual network of people, objects and spaces.

In contemporary societies, where economic value is generated through the processing of information and the monetization of knowledge, data in digital form has become the key factor in realizing social and economic goals. As modernity alienated *place*, *skill*, and *knowledge* to become "property", the transformation to the digital economy now reduces all these to a common numeric form to become data, e.g., *land* as digital maps, *skill* as factory automation and *knowledge* as inferred data, manipulated through computational machines.[2] This not only gives rise to the enclosure of information/knowledge in the commodity form of countable and exchangeable units (detached from any semantic content) but also extends datafication back over previous environments and other domains of life.

* * *

Much has been said and written about the data-driven practices and systems collecting vast amounts of data compiled from various sources and aggregated to obtain actionable information for purposes such as detecting fraudulent transactions, calculation of creditworthiness, organising Facebook newsfeed and so on.

Our digital *footprints* and *shadows* are either left behind unintentionally or delivered "willingly" as we conduct our daily affairs and situate our lives in various hybrid (physical and virtual) environments. We are all aware that we live with an ever-expanding *datafied* extension of ourselves[3], and the socio-economic order we are subjected to is heavily dependent on databases and computational tools to make decisions and implement rules of various kinds and scale. Due to the exponential increase and diffusion of data and the relevant computational tools, data-driven automated decision-making (ADM), deploying machine learning (ML) techniques, is becoming the "basis upon which the very fabric of the social world is being reconfigured."[4]

---

[1] Stephen Baker, *The Numerati*, Boston: Mariner Books, 2009, 29.
[2] Sean Cubitt, *Finite Media, Environmental Implications of Digital Technologies*, Durham: Duke University Press 2017, 159-163. Also *see* Karl Polanyi, *The great transformation: The political and economic origins of our time*. Boston: Beacon Press, [1944] 2001, Ch. 6 "The Self-Regulating Market and the Fictitious Commodities: Labor, Land, and Money".
[3] *See* Arnold Roosendaal*, Digital personae and profiles in law: Protecting individuals' rights in online contexts,* Oirschot:Wolf Legal Publishers, 2013.
[4] David Beer, "Productive measures: Culture and measurement in the context of everyday neoliberalism" *Big Data & Society*, January–June 2015:1–12, 1. https://doi.org/10.1177/2053951715578951

ML, a subfield of computer science and AI studies, is a way of designing and deploying algorithms which has evolved from the study of pattern recognition— using computation to discover useful regularities in data and exploit them to make predictions or select actions directly.[5] It is a general model of inductive learning in observational environments used for the analysis of large datasets. It invites and provokes types of empirical queries where the answer sought is not necessarily defined in advance.[6] ML accumulates a set of discovered dependencies, correlations or relationships that are referred to as "model". This inductive way of model construction is formalized and implemented through different learning methods.[7] These methods divide into subcategories such as *supervised*, *unsupervised, semi-supervised,* and *active learning.* This categorization is based on the way the data is represented, and other categorizations can be made based on the *goals* or *learning strategies*.[8]

The power of these systems and tools give rise to opportunities in various contexts of economic, social and political significance. However, these opaque, pervasive and intrusive technologies— viewed in a largely positive light as a valuable boost to efficiency— have direct influences on the social and economic relations of the individual and also on one's perception of self within society. From medicine to finance and immigration to criminal justice, automated decision-making (ADM) systems proliferating at a remarkable pace also imply radical changes to regulatory, administrative and managerial methods and procedures. ADM gives rise to concerns be it unfairness, discrimination, or arbitrariness with regard to eligibility/access to a specific service (e.g., bank loan, discounted flight ticket, etc.) or to certain benefits (social aid, health care, etc.).[9] ADM systems reach almost every realm of life including the gadgets we use and sensors that we encounter such as Internet of Things (IoT) and smart environments.[10]

Yet, more importantly, the ongoing replacement of human reasoning with computational processes (based on data analysis), primarily undermines the procedural safeguards which give flesh to the principles of due process and the rule of law.[11] This, at the outset, deprives individuals of any effective means to challenge the potentially unlawful, noncompliant, or unfair consequences of ADM— be it

---

[5] Joanna J. Bryson, "The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation" in Markus D. Dubber, Frank Pasquale, Sunit Das (eds), *The Oxford Handbook Of Ethics Of AI*, New York, NY: Oxford University Press, 2020, 2-25, 6.

[6] Valentina S. Harizanov et al., "Introduction to The Philosophy and Mathematics of Algorithmic Learning Theory" in M. Friend, N.B. Goethe and V.S. Harizanov (eds.), *Induction, Algorithmic Learning Theory, and Philosophy*, Dordrecht: Springer, 2007, 1-24, 2.

[7] Solon Barocas and Andrew Selbst, "Big Data's Disparate Impact" *California Law Review*, Vol. 104, 2016. http://ssrn.com/abstract=2477899

[8] Mehmed Kantardzic,. *Data Mining: Concepts, Models, Methods, and Algorithms*. 2nd ed. Hoboken, New Jersey: John Wiley, 2015, 89. " While unsupervised algorithms are not without their use in the legal domain, supervised algorithms are of the greatest saliency as they are driving the most legally consequential decisions in contexts such as risk assessment, immigration, predictive policing, credit scoring, and other contexts." Christopher Markou and Simon Deakin, "Ex Machina Lex: *Exploring the Limits of Legal Computability.*" in Simon Deakin and Christopher Markou (eds), *Is Law Computable: Critical Perspectives on Law and Artificial Intelligence*, Oxford: Hart Publishing, 2020, 31-66, 37 (footnotes omitted).

[9] As an example of the early literature reflecting these concerns, see Danielle Keats Citron, "Technological Due Process" (2008) *Washington University Law Review* 85(6):1249-1313.

[10] "[...] everything humans deliberately do has been altered by the digital revolution, as well as much of what we do unthinkingly. Often this alteration is in terms of how we can do what we do—for example, how we check the spelling of a document; book travel; recall when we last contacted a particular employee, client, or politician; plan our budgets; influence voters from other countries; decide what movie to watch; earn money from performing artistically; discover sexual or life partners; and so on." Bryson (n 5), 4.

[11] Margo Kaminski, "Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability" (2019) *Southern California Law Review* 92(6):1529–1616, 1545.

the breach of labour contract (e.g., data-driven performance assessment), violation of electoral laws (e.g., political micro-targeting) and so on. As such, the concerns relating to rule of law have wider connotations embracing a more general concept of procedural impairments where ML-based systems make up a substantial part of the decision-making process.[12]

Regarding these procedural impairments and the extent to which they affect decision-making process, ADM systems cannot be seen as a generic technology deployed uniformly. Systems like COMPAS[13] make use of ML in a rather static way, that is, once trained offline and deployed, the ML model does not go through a further learning process during its operation. In contrast, some more autonomous systems that are dynamically fed in real-time training data over time constantly adjust themselves (e.g., online trading systems).[14] In the practical domain, however, ADM systems do not precisely fit into this binary division but generally comprise of several ML-based tasks utilising different types of data in different ways. Depending on the intensity of data utilisation and the complexity of the system, the procedural impairments vary as to their effect and the contestability impediments they create.[15]

1.1. ADM: A question of contestability

The deployment of ML-based systems in contexts which have formerly required human judgment give rise to questions such as: *For what purposes do companies and governments utilize these systems? How accurate are the results and the underlying data used to create them? Who is deciding which criteria the decisions rely on? What are the legal or ethical basis of the classifications that these systems create and utilise? What are the privacy concerns? And more importantly, how could one challenge the outcome of these obscure systems and processes?* As seen, the use of algorithms for decision-making purposes could implicate several domains of law in various ways and this is also not limited to judicial or administrative domains but applies to all public and private sector decisions.

---

[12] "[...] algorithmic decision-making might be regulated differently or trigger regulation at different thresholds in different policy contexts and against the backdrop of different areas of the law." *ibid*. 1551.

[13] *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) is a case management and decision support tool developed and owned by Northpointe (now Equivant) used by U.S. courts to assess the likelihood of a defendant becoming a recidivist.

[14] "Algorithms have the potential to be highly dynamic, learning from new data as it becomes available. Or they can be relatively slow moving depending on when the responsible people get around to updating the system. [...] The temporal dynamics of algorithms create practical challenges for producing transparency information: What is the right sampling interval for monitoring and disclosure? To what extent should audit trails record internal and intermediate states of the machine? And how does this trade off against the resources needed for that monitoring? With algorithms potentially changing quickly, transparency presentations may also need to utilize dynamic or interactive techniques to convey information. This also raises the question of navigating and potentially comparing between different sets of transparency information." Nicholas Diakopoulos, "Accountability, Transparency, and Algorithms" in Dubber and others, (n 5), 197-213, 208. Art. 15(3) of AI Act proposal (see Conclusion Part 2.2) requires that AI systems that continue to learn after being put into use shall address the situations where biased outputs used as input for future operations ('feedback loops') with appropriate mitigation measures. For more on *static v. dynamic* ML, *see* https://developers.google.com/machine-learning/crash-course/static-vs-dynamic-training/video-lecture

[15] "In the broader domain of algorithms implemented in various areas of concern (such as search engines or credit scoring) machine learning algorithms may play either a central or a peripheral role and it is not always easy to tell which is the case. For example, a search engine request is algorithmically driven, however, search engine algorithms are not at their core 'machine learning' algorithms. Search engines employ machine learning algorithms for particular purposes, such as detecting ads or blatant search ranking manipulation and prioritizing search results based on the user's location". Jenna Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms", *Big Data & Society*, January–June 2016: 1–12.

Accordingly, increasing criticism and concern is witnessed about the transparency, oversight and the necessary legal mechanisms to remediate automated decisions that adversely affect individuals. The lack of transparency about the functioning and capacities of these systems is increasingly associated with risks for fundamental rights and procedural safeguards, including but not limited to rule of law. This has brought to attention the long-enduring provision of EU personal data protection regime on automated decisions (DPD Article 15 replaced by the GDPR Article 22).[16] The upgraded version of the right to object to decisions solely based on automated data processing in Article 22 of the GDPR has been the crux of the discussions about what transparency could mean within the context of ADM and how it could be implemented. The GDPR's framework on automated decision-making—with the newly introduced right to contest— is regarded to enhance key constitutional values by preserving human autonomy, increasing legal certainty and providing procedural safeguards.[17] The relevant provisions implement the constitutional value of due process in the context of algorithmic decision-making.[18] Through a series of rights, the Regulation translates the principle of the rule of law to the domain of ADM and private actors.

From a *contestability* perspective, what makes data-driven ADM systems problematic is that they rely on inferences or correlations drawn by algorithms. While decisions in the conventional sense are based on constructed rules (based on either normative, causal or logical grounds) and proven facts, ML facilitates the discovery of correlations between parameters. For example, online micro-credit services heavily rely on ML analysis of mobile phone data of the loan applicant. In this specific case, the battery charge level of the phone, charging periods and frequency and whether the phone gets turned-off due to running out of battery are a few of the thousands of parameters that one would have to scrutinise in order to challenge the decisions. Moreover, such scrutiny often reveals that an important part of the decisional criteria is legally or morally unacceptable and, in many cases, a larger sum inexplicable. In sum, regarding normative compliance or legality, in data driven ADM systems, the basis of the decision is unclear, and the outcome rather relies on a certain probability that one may behave (or may have behaved) in certain ways. The information asymmetry created by the insufficient transparency of these systems leaves us with a seemingly (if not actually) arbitrary decision.

In connection with the above, it should be mentioned that while some data driven ADM systems directly apply norms (e.g., monitoring violations/adherence, sanctioning), some simply operate in a normative framework which requires compliance with certain norms. In terms of normative contestation of an ADM process, both types of ADM systems present similar difficulties and thus give rise to similar procedural impairments. In the micro-credit example above, there is a contractual and/or legal regulatory framework that the system should adhere to. Where such mobile phone data is analysed by a public body to determine eligibility to social housing, the contestation problems would not

[16] Data-driven practices have been part of the emerging information society since the nineteenth century, and the attempts to regulate algorithmically controlled systems could be traced back to 1978 in Europe in Art.2 of the French code *Loi informatique et libértés* and to 1984 in the USA as the Congress passed the regulation "Display of Information, which required that each airline reservation system "shall provide to any person upon request the current criteria used in editing and ordering flights for the integrated displays and the weight given to each criterion and the specifications used by the system's programmers in constructing the algorithm. Matthias Spielkamp (ed.) *Automating Society Taking Stock of Automated Decision-Making in the EU*, 2019 AlgorithmWatch, Berlin, https://algorithmwatch.org/en/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf , 8.

[17] For such dimension of the GDPR generally *see* Kaminski (n 11), 1586-1595.

[18] Edoardo Celeste, Giovanni De Gregorio, "Digital Humanism: The Constitutional Message of the GDPR", *Global Privacy Law Review*, 2022, 3(1): 4-18.

radically differ from the micro-credit case. For another example, imagine an ADM system which is trained by visual data to apply traffic rules through video footage and accordingly sanction offenders. With some modification, the same ML model could also be implemented in a self-driving car to ensure compliance with the traffic rules. In this mode of application, the system could also embody further driving rules (e.g., prioritising reducing carbon emissions) based on the preferences of the driver or car owner. In these scenarios, both the decision in the form of a traffic fine and in the form of a certain driving behaviour will give rise similar transparency requirements for the effective contestation or scrutiny of the respective outcome. The difference in the latter case is that the contestation will come into question in connection with a compensation claim against the vehicle owner, car manufacturer or the developer of the self-driving system. Hence, the legal relevance of ADM systems should be understood in a broad sense, that is, even in the case of discretion of public authorities or freedom of contract granted to private parties, the decisions could still not be arbitrary and should adhere to certain public law principles and/or civil law limitations on private rights and contracts such as the *culpa in contrahendo*[19] doctrine. In modern societies, almost no part of human action or interaction is without legal or at least some type of regulatory context.[20] Moreover, considering the increasing legislative intervention in the area of data use, digital services and AI deployment, it becomes difficult to clearly discern the difference between rule application and rule compliance for some of the actors who develop or deploy these systems. This broad approach to legal relevance also aligns with Article 22 of the GDPR which targets any type of decision which has legal or similarly significant effects on the data subject.

1.2 Transparency in the context of contestation

Data-driven techniques mark a shift from theory-driven knowledge acquisition to a discovery-driven methodology.[21] Rather than starting with a question or theory, data-driven systems first deploy algorithms to look for patterns and correlations through which certain decisional input (similar to fact or evidence) is inferred.[22] These systems test several solutions or hypotheses for a given problem to choose the best fitting one.[23] This suggests that the expansion of data-driven technologies radically transforms what counts as *known (fact), probable and certain*.[24]. These systems do not simply record facts

---

[19] Civil law doctrine of *culpa in contrahendo* requires that contracting parties are under a duty, classified as contractual, to deal in good faith with each other during the negotiation stage, even in cases where parties fail to reach an agreement and no contract is concluded in the end. Friedrich Kessler and Edith Fine, "Culpa in Contrahendo, Bargaining in Good Faith, and Freedom of Contract: A Comparative Study." *Harvard Law Review* 1964, 77(3): 401–49. https://doi.org/10.2307/1339028.

[20] "All human activity, particularly commercial activity, occurs in the context of some sort of regulatory framework. The question is how to continue to optimize this framework in light of the changes in society and its capacities introduced by AI and ICT more generally." Bryson, (n 5), 9.

[21] Mireille Hildebrandt, "Defining Profiling: A New Type of Knowledge?" in Mireille Hildebrandt and Serge Gutwirth (eds) *Profiling the European Citizen*, Dordrecht:Springer, 2008, 17–45. http://www.springerlink.com/content/r70n22p620k62301/abstract/.

[22] David Chandler, "Digital Governance in the Anthropocene: The Rise of the Correlational Machine" in David Chandler and Christian Fuchs (eds.), *Digital Objects, Digital Subjects*, *Interdisciplinary Perspectives on Capitalism, Labour and Politics in the Age of Big Data*, London: University of Westminster Press, 2019, 23-42. https://doi.org/10.16997/book29.b

[23] Hubert Dreyfus, *Alchemy and Artificial Intelligence,* Rand Corporation, 1965. Also see Paul R. Thagard *Computational Philosophy of Science,* Cambridge, Massachusetts London*:* MIT Press,1988 (the first paperback ed. 1993), 157-173.

[24] "New technologies for automated surveillance and prediction neither simply augment human reason nor replace it with its machinic counterpart. Rather, they affect the underlying conditions for producing, validating, and accessing knowledge and modifying the rules of the game of how we know and what we can be expected to

about the *World* but transform data into a representation of "reality" which forms the basis of the decision-making process. This representation is an abstraction in the sense that certain properties and characteristics are ignored because they are regarded as peripheral or irrelevant to the task in hand. The ML model constructed through repeated observations over time and/or space does not necessarily explain but purports to rationalize what otherwise would be regarded as coincidental or unknowable.[25]

Accordingly, transparency for the purpose of contesting ML-based decisions takes on new meaning, requiring an understanding of the patterns or the rules that have been used to reach a decision. When ML-based data-driven systems are used for automating decisions, they are no longer a method of filtering data, but a way of outsourcing decision-making from human to machine or to software. Hence, a normative scrutiny of ADM entails a reconstruction of the process in a way akin to the decisions in a legal system or in other regulatory or normative framework— comprising of *facts*, *norms* and the ensuing consequences.[26] However, this is easier said than done simply because as these systems rely increasingly on data-driven inferences and correlations, the normative aspects of the process lose salience. The more complex the system becomes, the more the causal relations, goals, and intentions underlying the system become blurred and seem less relevant in terms of interpreting the decision.[27]

In addition to the technical and procedural difficulties in establishing such a multi-dimensional and versatile understanding of transparency, ADM systems are almost entirely shrouded in legal and commercial secrecy both in terms of the inferred models and the training data. On the legal front, demands for transparency have evoked counter-arguments mostly based on Intellectual Property (IP) rights and Trade Secret protection.[28] Consequently, in the event that a loan applicant wishes to examine and understand the software which calculates her credit score, the developers and operators of the credit scoring system would probably oppose this request on the grounds that the inner workings of the system are protected as *trade secret* and thus may not be disclosed. So, we are faced with a modern clash of rights between the rights of individuals and businesses which aim at harnessing data as an asset. As transparency relates to communication of certain knowledge or information, the link with the rules governing creation, dissemination, and use of knowledge is evident. One part of this knowledge relates to governance of data (personal or otherwise), the other part relates to IP rights where this knowledge takes the form of creative works (copyrights), industrial processes (patents), or hybrid

know." Sun-ha Hong, *Technologies of Speculation: The Limits of Knowledge in a Data-Driven Society*. New York: New York University Press, 2020, 1-2.

[25] Adam Jacobs, "The Pathologies of Big Data", *Communications of the ACM,* 2009, 52: 36-44.

[26] "Legal systems are sets of rules for what is allowed, frameworks for what rights people have, and plans for what kind of society we will live in. Technical systems do the same things in different ways. They are sets of rules for what is (not) allowed, frameworks for what rights people (don't) have, and plans for what kind of society we will (not) live in. Technologies are like legislation: there's a lot of them, they don't all do the same thing, and some are more significant, but together as a system they form the foundation of society." Jathan Sadowski, *Too Smart: How Digital Capitalism Is Extracting Data, Controlling Our Lives, and Taking over the World*, Cambridge, Massachusetts: MIT Press, 2020, 6. As Markou and Deakin put, "[b]ecause much of legal reasoning is algorithmic, there is huge scope for ML applications in the legal context." Markou and Deakin (n 8).

[27] "For example, in situations where variables are confounded, it can be challenging to establish whether a measured effect is causal or illusory. Confounding occurs when multiple factors correlate with a certain outcome, and there is confusion over which associations represent the cause, limiting the extent to which any one can be assigned responsibility." Joshua Kroll, "Accountability in Computer Systems" in Dubber and others (n 5), 189.

[28] "[f]ull transparency often trades off with other values related to confidentiality. Whether confidentiality protects the personal privacy of individuals affected by a computer system or the proprietary intellectual property interests of the system's creators or operators, the level of transparency required for governance often trades off the disclosure of legitimate secrets." *ibid*. 194.

elements such as software or databases. Given that both the training and the deployment of ADM systems rely on processing several types of information/data gathered from a multitude of sources— where multiple parties can claim conflicting rights— questions about the legal entitlements in data become ever more complex and all the more important.

The bottom line is that the implementation of transparency for the purpose of contesting automated decisions requires several regulatory and procedural mechanisms, as well as technical tools to provide proper insight into ADM systems in various dimensions.[29] This also includes the main assumptions and hypotheses underlying the ML model which are directly linked to the purpose(s) for which the ADM has been deployed.[30] As such, transparency in the sense of legibility is an essential ingredient to ensure that ADM, both in the private and public domain, complies with certain procedural due process requirements.[31] It is deeply rooted in the adversarial principle and fully resonates with the procedural safeguards, e.g., the right to raise an argument about the evidence and about the norms relevant to the decision. As Waldron put it[32] :

> "Applying a norm to a human individual is not like deciding what to do about a rabid animal or a dilapidated house. It involves paying attention to a point of view and respecting the personality of the entity one is dealing with. As such it embodies a crucial dignitarian idea— respecting the dignity of those to whom the norms are applied as *beings capable of explaining themselves*."

2. The objectives of the study

Taking stock of this background, the thesis initially puts forward the perspective that transparency in ML-based decisions is not about reading of computer code but rather relates to the question how these systems make up the regulatory realm that we are subjected to. The theoretical basis of the research is to approach data processing and ADM as regulatory technologies. So, as a preliminary premise, the thesis treats data protection law as a kind of 'meta-level regulation' —a body of rules, regulating a regulatory technology. As an extension of Lessig's "Code as Law", the thesis rests on the premise that by *sorting, classifying* and *predicting* these technologies impose or facilitate certain norms, values or criteria.[33] In comparison to early expert systems, current data-driven technologies—constantly fed by

---

[29] The 'Ethics Guidelines on Trustworthy AI' specify that in case of a 'black box' obstacle, alternative measures such as traceability, auditability, and transparent communication on the capabilities of the AI system should be considered. The required degree of transparency is highly dependent on the context and the severity of the consequences where the output is erroneous or otherwise inaccurate. European Commission, High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (2019).

[30] "Though justice and legal certainty require equal treatment, they cannot provide the measure or nature of the treatment, for which we need an understanding of the purpose of the treatment, and a decision" Mireille Hildebrandt, "Radbruch's Rechtsstaat and Schmitt's Legal Order: Legalism, Legality, and the Institution of Law." *Critical Analysis of Law,* 2015, 2:1, 52.

[31] Ida Varošanec, "On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI", *International Review of Law, Computers & Technology*, 2022, 36(2): 95-117

[32] Jeremy Waldron, "The Rule of Law and the Importance of Procedure", NYU School of Law, Public Law Research Paper, 2010, No. 10-73, 14. https://ssrn.com/abstract=1688491.

[33] "Algorithms are one of the frontiers of normativity. While we spend an increasing part of our lives interacting with digital devices and online platforms, software is becoming the de facto regulator of human societies." Nicola Lettieri, "Law in the turing's cathedral: Notes on the algorithmic turn of the legal universe." in Woodrow Barfield (ed.) *The Cambridge Handbook of the Law of Algorithms* United Kingdom; New York, NY: Cambridge University Press, 2020, 691-721, 701. Also see, Bert-Jaap Koops, 'Criteria for Normative Technology: An essay on the acceptability of 'code as law' in light of democratic and constitutional values'

data collected from sensors, personal devices, social media, and online transaction— are claimed to enable more granular, "effective" and instantaneous command over things, people, and places.

Based on the premise that transparency can empower individuals to make informed choices and to judge the potential consequences of emerging technologies, a significant part of the scholarly research and the ensuing debates have focussed on the outcome that the concept of transparency targets (e.g., an explanation) but not on the possible forms and constructs of transparency or the specific mechanisms that contesting automated decisions would entail.[34] Accordingly, the two main objectives of the thesis are:

- To define the essentials of a transparency model (contestation scheme) which is based on a theorisation of ADM as a techno-regulatory process where certain input leads to certain results. The idea is to explore what interpreting the algorithm could mean (other than explanation) for the purpose of contesting automated decisions. The analysis intends to shift the focus from *how to understand the algorithm* to an earlier question: *how it should be understood.*
- Developing a systemisation and a methodological analysis of the relevant provisions of the GDPR—focussing on the practical implementation of the remedies provided under Article 22 of the GDPR (i.e., rights to *obtain human intervention*, *to express one's views* and to *contest* the decision). By addressing the specific transparency implications of the "right to contest", the thesis aims to transcend the current debates about the existence or the scope of a so-called "right to an explanation". Put differently, unlike a significant part of the literature, the analysis in the below chapters is not mainly concerned with the outcome of ADM as might be unfair, discriminatory, or biased in general, but exploring the methodologies and practical requirements enabling effective contestation.

The proposed approach to transparency is an overarching framework both intended (1) as a guidance for the design and audit of ADM systems, and (2) as a *scheme* for the ex-post scrutiny of specific decisions. Along with this, (3) achieving terminological consistency and structural robustness within a conceptual framework has also been one of the major concerns of the study.

3. The major research question and sub-questions

Having identified the above legal challenges that are confronted when algorithms are used for decisions which traditionally require human judgment, the research question (RQ) on the conundrum of transparency in ADM takes shape as:

How could transparency of automated decision-making be constructed as a prerequisite of contestation and what could be the main affordances and impediments under EU law?

The thesis is focused on a specific type of transparency, that is, an actionable one for the purposes of contestation. Given that the contemplated research is primarily concerned with how ADM systems must be constructed and administered, the first sub-question (first paper) materialises as:

[34] On different regulatory approaches to ADM and whether individual due process rights can address the concerns, see Kaminski (n 11) 1553ff. and for an account of the current literature fn. 118.

(1) what types of impairments or harms do automated decisions (as regulatory processes) give rise to in terms of contestation? In the simplest terms, why do we need transparency?

As the harms relevant to contestation become formulated as *rule of law* implications (a.k.a. procedural impairments), the second sub-question means to understand (second paper):

(2) What are the transparency challenges stemming from the procedural impairments (rule of law implications) and how could they be systemised?

As the findings of this analysis clarify the relevant aspects of the notion of transparency from the contestation perspective, the *third* sub-question (second paper) focuses on the solution and seeks to answer:

(3) what are the essential requirements of a transparency model to contest automated decisions?

Next, the thesis proceeds with the analysis of positive law. As the most advanced and relevant legal framework, the *fourth* sub-question (third paper) inquires:

(4) to what extent is the current EU data protection regime (GDPR) compatible with the transparency desiderata laid out by the answer to the third question?

The *fifth* sub-question (fourth paper) is devoted to possible impediments under the EU IP regime, exploring:

(5) to what extent could IP rights function as a barrier against the transparency demands that aim to render automated decisions contestable?

As such, the thesis does not intend to offer a full-scale analysis of all legal implications of transparency in ADM systems but rather concentrates on certain contestation-specific problems in relation to two prominent legal regimes, namely personal data protection and IP rights.

It should be specifically noted that the main objective of the thesis is not to discuss the general framework of automated decisions under EU law, but rather to see whether the relevant provisions of the GDPR and the IP regime could accommodate the transparency model and the implementation mechanisms defined and systemised by the thesis. In line with this, other current and upcoming legislation relating to automated data processing or decision-making have been left out of the scope of the thesis[35]. The reason that the thesis focuses on Article 22 and other transparency-related provisions of the GDPR is because the Regulation is the single comprehensive framework in the EU *acquis* covering both public and private sector decisions which have legal or similarly significant effects (with

[35] Despite its close connection and similarity with the GDPR, the thesis also does not examine the Directive EU 2016/680 (Police Directive). Even though the Directive contains important EU data protection instruments that regulate automated individual decision-making, the relevant Article 11 does not allow the data subjects to express their point of view or contest the decision. As this omission is a deliberate choice of the legislature, the question then centres around whether a *right to contest* could be implied from the *right to obtain human intervention* in Article 11(1) and the general rules of procedure. The matter requires an extensive analysis of the Police Directive, especially a thorough consideration of the security concerns. Yet, it could simply be mentioned that if the involvement of human intervention or review by non-automated means is transparent and not concealed from the data subject, the decision may inevitably be subject to contestation, at least on the grounds of administrative arbitrariness or malintent. For other EU legislation briefly evaluated in the Conclusion, *see* Ch.6, sec.2.2.

a detailed transparency scheme and several provisions supporting various technical and institutional measures as elaborated by the thesis). Nevertheless, the transparency model of the thesis could offer guidance for the implementation of other legislation providing for the scrutiny or review of ADM systems. Additionally, it can contribute to the development of a systemic approach to various transparency-related provisions found across numerous legal documents. The abstract and domain-agnostic nature of the model allows for flexibility in adapting it according to the specifities of the targeted ADM system.

The relevance of the thesis lies in the fact that both the developers and users of ADM systems—aiming to exploit data either obtained from open sources or through contractual acquisitions— have to address a number of legal conundrums primarily stemming from IP and DP Law. Hence, the initial decisions about the selection of the data types and the ML models play a decisive role in the lawfulness and legal compliance of the final service or product. Similar legal challenges also restrict those who wish to challenge the outcome of ADM systems as well as those who develop tools for this purpose. As such, the thesis is also intended to serve as a methodological guideline both for individuals and for the relevant industries while offering input for the policymakers who are in need of a thorough picture of the legal landscape on this particular matter.

4. Methodology

Although transparency is a legal concept, it is not legally defined, or its essential characters cannot be precisely determined purely by reference to legal norms. This entails an interdisciplinary and accordingly a pluralist approach to the methodology of the contemplated research. The interdisciplinary character stems from the fact that the intended analysis first, needs to conceptualise a specific technological application, namely data-driven ADM, and then to define the relevant transparency requirements which may also be articulated as a *model* in the broadest sense.

Scientific disciplines develop theories to discover, explore, and control phenomena and to systematise, organise, and summarise the knowledge accumulated by them. For this thesis model building, and conceptualisation (concept formation) are closely intertwined in that these efforts define and systemise diverse elements relating to transparency for the intended analysis. All these methodological tools imply abstraction, logical coherence and, simplicity to the extent possible.[36] Theorisation and model building together with the ensuing typologies and systematisations are the primary methods of the thesis to break down complex, multifaceted phenomena such as ADM into manageable parts.

By conceptualising ADM as a type of techno-regulation and subsequently defining transparency requirements (aiming for contestation), the first two papers employ a hybrid methodology which is not primarily rooted in the legal discipline. Deriving from a rich literature relating to scrutiny and transparency of data-driven systems in various domains such as STS, philosophy of technology and computation, epistemology, communication studies as well as managerial and regulatory sciences, the theorisation (conceptualisation) and the model building aim to situate the complex and amorphous concept of transparency in a context which is amenable to legal analysis.

[36] Mark van Hoecke, "Legal Doctrine: Which Method(s) for What Kind of Discipline?" in Mark van Hoecke, ed. *Methodologies of Legal Research: Which Kind of Method for What Kind of Discipline?* Oxford ; Portland, Or: Hart, 2011, 1-19, 16.

The thesis rests on a theorisation/conceptualisation of ADM as a techno-regulatory process and accordingly a modelling of transparency from the perspective of contestation. These theoretical efforts aim for a system of coherent and non-contradictory assertions and concepts, that enable to identify different forms and constructs of transparency relevant to the scrutiny of ML-based ADM. To properly weave the conceptual links between transparency and contestation in ADM, the development of a *contestation-specific* typology of transparency challenges (informational asymmetries) is among the primary tasks in the thesis and is expected to contribute to the terminological arsenal to conceptualise ADM systems at a sufficient level of abstraction and thus, enable the analysis of a diverse array of legal implications relating to transparency in ADM.

This theoretical perspective informs us about what we want to know, how we could obtain the answers and thus paves the way for an inquiry to address various types of obscurities in ML-based decisions. Yet, any theorization, systemization or model-building and the ensuing interpretation inevitably relies upon a legitimation of a shared world view, common basic values and norms— meaning that these methodologies are rather justificatory in nature.[37] Such inquiry inevitably requires broadening the scope of analysis to an interdisciplinary direction, extending to various social and technical domains. As a methodological difficulty, such endeavour raises the question how far a legal researcher could reach without exceeding his/her limit of competence. Considering these challenges, the thesis takes contestation as an anchoring concept and approaches ADM as mechanisms with regulatory consequences. The thesis addresses transparency and contestation as a "procedural" problem in the sense that these concepts primarily relate to *legal reasoning and adjudication* rather than the substance of the rights. This enables an analysis of ADM systems not by their inner workings but rather by the implicit "normativity" embedded in their behaviour/action. Hence, an assessment of the consequences of ADM as to their lawfulness under the fundamental rights and constitutional principles (e.g., autonomy, privacy, equal treatment, freedom of speech) is not of relevance to the perspective of this study.

In terms of decision-making process, ML-based systems mark a shift from *causation to effects*.[38] Approaching ADM as a regulatory process (and consequently transparency from a procedural perspective) may be seen as an antidote to solutionist tendencies which are primarily concerned with the effects of the decision (fair, equal, just etc.)[39] and hence *unbiasing* the technology[40]— neglecting the fact-gathering and reasoning processes as they are much less comprehensible from the outside.[41]

---

[37] For arguments that legal interpretation is inherently normative, see Anne Ruth Mackor, "Explanatory Non-Normative Legal Doctrine. Taking the Distinction between Theoretical and Practical Reason Seriously" in van Hoecke, (n 36), 45-70, 58; Aleksander Peczenik, "Scientia Juris. Legal Doctrine as Knowledge of Law and as a Source of Law" in Enrico Pattaro (ed), *A Treatise of Legal Philosophy and General Jurisprudence*, Dordrecht: Springer, 2005.

[38] Chandler (n 22), 24. Also see Rostam J. Neuwirth, *Law in the Time of Oxymora: A Synesthesia of Language, Logic and Law*. Juris Diversitas. London ; New York, NY: Routledge Taylor and Francis Group, 2018, 3-4.

[39] As Hildebrandt puts: "Equal treatment will have a different meaning when law serves the goals of liberalism, than when it serves the goals of collective well-being".She further notes that according to Radbruch "[w]hile justice directs us to treat equals equally, unequals unequally, it does not tell us anything about the viewpoint from which they are to be deemed equals or unequals in the first place; moreover, it determines solely the relation, and not the kind, of the treatment." Gustav Radbruch, "Legal Philosophy" In: *The Legal Philosophies of Lask, Radbruch and Dabin* (Edwin W. Patterson ed., Kurt Wilk trans.) Cambridge, Massachusetts: Harvard University Press, 1950, 107 cited in Hildebrandt (n 30), 50.

[40] Matthew Le Bui and Safiya Umoja Noble, "We're Missing a Moral Framework of Justice in Artificial Intelligence: On the Limits, Failings, and Ethics of Fairness" in Dubber and others, (n 5), 166.

[41] Rónán Kennedy, "The Rule of Law and Algorithmic Governance." in Woodrow (n.33), 209–326, 228.

Inquiries which are mainly result-oriented are likely to overlook that ADM can be deployed in a bewildering array of ways that are often difficult to pin down and map out in their entirety. This is especially the case for this study since it aims for a broad scope of automated decisions which deploy various types of ML techniques in varying intensity, and which relate to law or any other normative framework in a wide spectrum including both norm application (enforcement) and norm compliance.

* * *

As the first two papers of the thesis define the problem space, the following papers employ hermeneutics to provide an interpretation of the regulatory framework and to test the consistency of this interpretation with the transparency model provided in the *second* paper. Accordingly, the main research subject of the *third* and *fourth* papers are the relevant provisions of the personal data protection and IP regimes under the EU *acquis*. Both papers, respectively, engage in a systemisation of the transparency-related provisions of the GDPR and the ML elements from an IP perspective.

Interpretation, as the core business of legal doctrine since the Roman times, is deeply entangled with legal practice and contains both evaluative and descriptive aspects. It reflects the argumentative and adversarial character of law. Interpretation and argumentation are tightly intertwined in that (both in legal doctrine and in legal practice) each interpretation needs arguments when diverging interpretations could reasonably be sustained. Taking a step back from the interpreted text, the argumentative process produces answers to concrete legal questions based on values, goals and higher principles and constitutional caveats that are relevant in a given context. Interpretation and argumentation may be seen as two sides of the same activity—while interpretation is the goal, argumentation is the means for sustaining the interpretation.[42]

Legal argumentation and reasoning are not only aimed at finding the contents or imperatives of the law but also convincing one's auditorium of a particular legal position.[43] This reveals the complementary and dialectical relation between theory and interpretation. Determining the exact meaning and the scope of the transparency-related provisions of the GDPR also requires contemplation of the validity and the precise meaning of a legally relevant text. This inevitably depends on the outcome of prior conceptualisations and theorisations, especially when the result of a literal interpretation leads to unreasonable or incompatible results.[44]

5. Outline of the chapters

5.1 Procedural implications— why the need for transparency?

The first paper of the thesis, co-authored with Ronald Leenes, *'Rule of Law' implications of data-driven decision-making: a techno-regulatory perspective,* prepares the ground by conceptualising data-driven ADM as the new horizon of techno-regulation. Although the paper directly refers to rule of law as a terminological choice, it targets a broader context to explore ADM systems from a due process perspective. In this respect, the analysis of the paper is not confined to legal rules but generally concerns

[42] van Hoecke (n 36), 5.
[43] Most arguments in legal reasoning are not 'true' or 'false' but more or less convincing. They do not qualify for an empirical verification. *ibid*.
[44] *ibid*. 14.

any type of scrutiny where the decision in question relies upon or is justified on normative grounds, be it legislation, contract, ethical principles, code of conduct and so on. As such, the analysis identifies three types of harms (due process/procedural impairments) which may be extrapolated to a wider realm of data-driven automated decisions.

The *first* impairment is the *collapse of the normative dimension*. This refers to how the normative dimension of law or the regulatory framework is replaced with patterns in datasets when decisions of legal significance are entrusted to data-driven systems. Moreover, as mentioned above, the normativity imposed by ADM systems is not stable, but rather emerges from the data used for training the system. Especially in certain types of ADM systems which directly deploy ML to produce results, what is regarded to be the 'norm' is not necessarily predetermined.[45]

Second, ADM gives rise to an impairment as to the causal (logical) dimension of the decision-making, process— drawing attention to the difficulty in distinguishing between the events that are causally or logically related and those that are merely correlative.[46] Simply put, data cannot tell us whether people ate ice cream when summer came or vice-versa, or whether the sun will still rise tomorrow even if the rooster were silent. Patterns appearing in large datasets may be nothing more than spurious relationships. This severely impedes the individual's capacity to scrutinise data-driven decisions.

For the *last* impairment, we may speak of the demise of Law as a *moral enterprise*. The impediments to contestation caused by causal and normative impairments also imply that there exists no authority or agency to justify and accordingly take the moral responsibility of the outcome.[47] Argumentation and adjudication not only provide redress, but also have a connotation of morality that justifies the outcome and renders it acceptable. The idea of ML-based regulation, which hinders argumentation, discards this moral signalling dimension of law.[48]

5.2 The challenges and the consequent "transparency model"

The *second paper* provides a comprehensive evaluation and synthesis of the implications of 'transparency' within the context of contestation. Initially, under a threefold approach, the paper explains how i) intransparencies and opacities; ii) epistemological flaws (spurious correlation or weak causation); and

---

[45] "This line of questioning leads us to an even more productive question: what is a "norm" in the algorithmic world? In most instances, it's like Google's 'man' or Quantcast's 'Hispanic.'…Instead of abiding by a hard-coded fixity, the category of 'woman' quietly changes, strategically locating new key elements for its 'gender' while abandoning those that are ineffective. These relentlessly recalibrating containers are how we are measured, talked about, and represented on the fly" John Cheney-Lippold, *We Are Data: Algorithms and the Making of Our Digital Selves*, New York: NYU Press, 2017, 137, 143.

[46] "Just as the relation between, for example, the heating of a gas and the expansion of the gas is not created but discovered by natural scientists, legal scholars do not create but discover the relation between, for example, committing a tort and paying compensation." Anne Ruth Mackor, " Explanatory Non-Normative Legal Doctrine. Taking the Distinction between Theoretical and Practical Reason Seriously', in van Hoecke, (ed) (n 36), 45-70, 53.

[47] "[...] causal responsibility, moral responsibility requires *agency*, or the ability to have behaved differently in a situation where control of the operative outcome could have been effected." Kroll (n.27), 190. For justificatory concerns in ADM, *see* Kaminski (n 11), 1553-1557.

[48] "Many people think the purpose of the law is to compensate, and obviously if we allow a machine to own property or at least wealth then it could in some sense compensate for its errors or misfortune. However, the law is really primarily designed to maintain social order by dissuading people from doing wrong." Bryson in Dubber and others (eds) (n 5), 13.

iii) biased processes inherent in ML create obstacles in terms of understanding and contesting automated decisions. Each of these informational asymmetries renders individuals prone to manipulation and potentially incapacitates them from appealing against the results that are—depending on the context—unlawful, noncompliant, unethical, socially problematic or somehow worthy of scrutiny. As an answer to the second research question, this part of the thesis provides a systematic typology of transparency challenges giving rise to procedural impairments. The analysis reveals that the transparency implications of ADM systems are too complex to be dealt with by addressing certain opacities or invisibilities. Transparency with a view to see the entire system 'at work', is an ever-expanding territory, that is, as you open black-boxes, you may just find more black-boxes. Instead, *the* outcome of ML-based systems may eventually be attributed to the values and assumptions that underlie the response of the system to a given input.

Accordingly, the rest of the paper explains the essential components of a transparency model (contestation scheme) that are formulated as: i) the data as "decisional input"; ii) the decisional rules contained in the system; iii) the context and further implications of the decision[49]; iv) the accountable actors (the third research question). *First*, regarding decisional input, as the initial step of contestation one needs to understand how the system gathers factual input about the phenomena under analysis. *Second*, comes the issue of normativity (the decisional rules contained in the system). Taken in the broadest sense, every decision— be it judicial, administrative or private— can be decomposed to discover which rules have been followed in what order.[50] However, in case of ML-based systems, the normative bases that underlie the decision do not reveal themselves easily. Think of an ADM system which considers spelling mistakes for predicting overweight individuals, in this case, the assumed causal or logical connection between the input and the result may not be explainable or justifiable, being only a spurious correlation. The *third* dimension which is indispensable for contestation, is the *actual impact* and the *context* of the decision. This primarily involves informing of the individuals about where the decision starts and ends, and whether the system interoperates with other data processing operations. For example, a simple credit score is not only used to decide whether one would get a loan, but it may also determine the loan pricing, the type of loan monitoring, and the amount of credit. The actual impact of the decision is further determined by the context, that is, the particular situation, environment or domain in which the decision is made. Contesting a decision may require not only the knowledge of why a decision was made, but also why a different decision was not made. *Lastly*, for effective contestation, individuals must be aware of the responsible actors, whom to appeal to and eventually whom to hold accountable.

5.3 The affordances of transparency under the GDPR

Having identified the essential transparency requirements (input data, decisional rules, the *actual impact and* context, responsible actors) for effective contestation, the third paper "*The right to contest automated decisions under the GDPR: (Beyond the so-called "right to explanation*") inquires to what extent the EU data protection regime, namely the GDPR, could properly accommodate different transparency

[49] On the context of the decision, *see* Andrew D. Selbst, "A Mild Defense of Our New Machine Overlords" (2017), 70 *Vanderbilt Law Review En Banc*, 87-105. https://ssrn.com/abstract=2941078

[50] The thesis takes the concept of *decision* and *decision-making* in an extent not limited to legal or regulatory domain in the strict sense. In that regard, a recommendation for medical diagnosis could also be understood as decision (with or without legal effect) which may be contestable on normative— not necessarily legal— grounds as defined in the *second paper*.

modalities and the relevant implementation tools aiming for contestability (the fourth research question).

The paper identifies and analyses the relevant provisions of the GDPR to further explore the compatibility of these provisions with the transparency requirements outlined in the second paper. It first situates Article 22 as the core remedy with an essential role in determining the contours of transparency in relation to automated decisions under the GDPR. This part provides a systematic and teleological interpretation of Article 22 of the GDPR on automated decisions, focusing on how *the right to obtain human intervention*, *to express one's views* and *to contest the decision* could practically be implemented. Among those, the right to contest is regarded as the backbone provision with a key role in determining the scope of algorithmic transparency under the GDPR.

Next, the provisions of the GDPR are analysed in a "two-layered" approach. The *first layer* (human-intelligible transparency) deals with access and notification, formulated as individual rights in Articles 13 and 14. These articles provide that data subjects shall be given information about the purposes of the processing for which the personal data are intended together with the legal basis for processing. Such information enables the "reverse-engineering" of the decision-making process with a view to understand the underlying normative set-up. The purposes pursued by the system are of direct relevance to the context of the decision. In terms of exercising the right to contest, the thesis identifies *purpose limitation, compatible use*, and *notification of the intended purposes* as important leverages, the implementation of which are dependent on certain enacted and applied transparency requirements.

Given that the transparency requirements of automated decisions could not be limited to access and notification (explanation) in the conventional sense, the 2nd layer consists of further administrative and technical measures that are, *design choices* that facilitate interpretability, *institutional oversight* mechanisms and *algorithmic scrutiny*. As many of these measures are directly linked to the concept of *Data Protection by Design and by Default* (DPbD) as provided in Article 25, the paper argues that the Article 25 is not limited to implementation of certain data-protection principles but may be extended to a notion of *"Contestability by Design"* (CbD). The paper further identifies ~~a~~ number of GDPR provisions which either explicitly provide *institutional, administrative* and *procedural* measures or allow for their implementation.

Finally, four implementation modalities (regulatory options) are introduced, combining *1st and 2nd layer* transparency measures to implement Article 22 based on the risks created by a specific application of ADM. The first option (category) would be banning certain type of ADM systems due to the enormity of the risks they create. A second modality is to permit ADM but to subject it to ex-ante design and procedural requirements accompanied with ex-post algorithmic scrutiny where necessary. A lighter modality of implementation is to allow ADM with only *ex-post* algorithmic scrutiny requirements. At last, where there exist no considerable risks, ADM could be permitted without any restrictions.

5.4 The relevance of IP rights

As the most important legal framework that could impede transparency efforts, the *fourth* paper, *ML and the relevance of IP rights: An account of transparency requirements for AI,* provides a macro-view of the potential areas of conflict between the transparency requirements implicit in the right to contest and the relevant IP regimes—i.e., copyright, sui generis database right and trade secret protection (the fifth research question). The paper initially clarifies that the implementation of transparency measures and

mechanisms (e.g., the algorithmic audit or black-box testing of these systems) may require the disclosure, reproduction or modification of certain informational elements of ML systems.

This brings up the questions: i) which specific types of transparency implementation could give rise to IP infringement claims? ii) taking into account the statutory exceptions and limitations restricting IP rights, how could these IP claims be legally confronted from the vantage point of transparency? and iii) what may be the possible solutions within the IP regime?[51]

In order to properly address these questions, Section II of the paper lays out the structure of the IP analysis which examines ML systems in a dual ontology as: data and datasets (expressional elements[52]) on one side and *algorithmic techniques* and *ML models* (operational elements) on the other.[53] Next, Section III focuses on copyright and *sui generis* database protection of data and databases in a tripartite structure as: i) the training and test data, ii) the "actual" data analysed for a specific decision, iii) data comprising of ML output (predictions, ratings etc). It is examined under which conditions *extraction from* or *transformation of* databases to scrutinise ML systems could amount to an infringement of the *sui generis* database right. The analysis further extends to whether machine generated data could pass the substantial investment test regarding the obtaining, verification or presentation of the contents of the database as provided by the Database Directive. This part further includes an assessment of how the *text and data mining* (TDM) exception in the DSM Directive could facilitate the transparency demands.[54] Considering their relevance, this section also explores European Commission's recent legislative initiatives (i.e., Data Act[55] (proposal) and Data Governance Act[56]) aiming to foster data economy as laid out in European Data Strategy (2020).[57] These two proposals introduce various provisions which intend to limit either the substance or the exercise of certain IP rights for the sake of facilitating access to data.

Section IV deals with the IP protection of a*lgorithms (algorithmic techniques)*, *ML models* and the implementing *computer code (*as the operational/functional elements of ADM). The analysis is based on the view that the transparency mechanisms such as the software tools for audit and testing may necessitate the reverse engineering or the implementation (thus reproduction) of the computer code or the essential parts of the system. The resulting IP implications are discussed in relation to both

[51] The analysis of the paper does not extend to a discussion of the problem within the context of possible conflict between the fundamental rights (personal data protection and property) which requires the application of the proportionality test as laid out by the CJEU judgments. This dimension of the problem is briefly discussed in the conclusion.
[52] The term *expressional elements*, as used in this paper, refers to the creative or literary dimension of a copyrighted work as opposed to its conceivable function.
[53] As an exception to this structuring, trade secret protection of both the operational and expressional elements are dealt in a separate section. This is for the reason that trade secret law is not confined to a specific type of intellectual labour (e.g., artistic/literary work, database or technical invention) but protect any type of information against unlawful appropriation.
[54] Articles 3 and 4 of the Council Directive 2019/790/EC of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Council Directives (EC) 96/9 and 2001/29 [2019] OJ L 130 ('DSM Directive').
[55] Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM/2022/68 final.
[56] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).
[57] European Commission, "A European Strategy for Data" (Communication) COM(2020) 66 Final. The Strategy document envisages a series of legislation which will lift the barriers and increase the availability of data especially for small and medium enterprises.

copyright (granted for the literary elements of computer programs) and patent law (where ML-based systems, or parts of them, qualify as novel technical inventions).

The final part of the paper inquires how *trade secret* protection could act as a barrier against transparency demands. As confidentiality is a highly appealing mechanism for securing ML-based systems, in many cases, relying on trade secrets is the preferred legal strategy of the designers and operators of ADM systems. The analysis clarifies that both the *expressional* and the *operational* elements of ML systems could satisfy the *secrecy requirement* under the EU Trade Secret Directive.

The paper concludes that the existing exceptions and limitations of the European IP regime may be interpreted to permit the use of data and ML tools (both by the affected individuals and the supervisory authorities) for the purposes of algorithmic transparency. However, such progressive interpretation, albeit theoretically possible, seems to be fraught with many challenges at the practical and judicial level.

5.5 Findings, conclusions, and the temporal framing

The conclusion of the thesis evaluates the output in two parts, handling transparency in ADM and IP implications separately. Extending beyond the specific findings provided in each paper of the thesis, the conclusionary part on transparency rather aims to contextualise the compiled papers within the wider perspective of entanglements between law and data-driven processes.

The part of the conclusion on IP protection refines the finding that it is rather the rights which address digitalisation (e.g., software protection, sui generis database right) together with trade secret protection that mainly create barriers in the implementation of transparency. As the IP analysis in the *fourth* paper is restricted to the potential areas of conflict between the IP regime and the transparency requirements, the question of conflict between the fundamental rights to property and data protection has been left untouched. The conclusion briefly returns to this question illustrating arguments about the application of the proportionality test of the CJEU and emphasizing that, where trade secrets are at stake, the requirement of physical secrecy makes the assessment of a balancing between the rights difficult.

Regarding the temporal framing, the thesis spanned a period of nearly eight years from September 2015 to January 2023 as the closing date. The *first* and the *second* paper have been published in 2018 in March and December respectively. Both the introduction and the conclusion intend to cover certain literature that became available subsequent to the publication of these papers.[58] The *third* paper covers the legislative, judicial, and policy developments until March 2021, and the *fourth* paper until April 2022. Considering the temporal gap between the publication of the papers and the closing date of the thesis, the conclusion also examines the European Commission's newly enacted and upcoming legislative proposals as well as the judgment of the ECJ in *Ligue des droits humains* (C-817/19).

Due to its relevance regarding the implementation of transparency and the scrutiny of ADM systems, the conclusion first examines the European Commission's legislative proposal, AI Act[59] The proposal, which was released on 21.04.2021, aims to ensure responsible deployment of AI technologies while addressing the risks for fundamental rights and laying down harmonised transparency rules for certain

[58] The *first* and *second* paper of the thesis do not contain any substantial judicial or legislative references.
[59] Proposal for a Regulation of the European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act).

AI systems. As explained in the conclusion, although the Act does not contain individual rights or remedies that enable contestation of AI-based decisions, it provides substantial support to distinguish and crystallise the matters of transparency and contestation in a procedural context.[60]

Other ADM and transparency-related legislation examined in the conclusion are the regulations Digital Services Act (DSA)[61] and Digital Markets Act (DMA)[62]. These two regulations (which came into force subsequent to the publication of the relevant paper) form a single set of rules aiming to: i) create a safer digital space in which the fundamental rights of all users of digital services are protected; ii) establish a level playing field to foster innovation, growth, and competitiveness, both in the European Single Market and globally. This part provides a limited review of these legislation which contain transparency provisions relating to recommender systems, content moderation and in general profiling practices of online platforms.[63]

As a significant judicial development, the conclusion further looks into the decision of the European Court of Justice (ECJ in the case *Ligue des droits humains (LDH).*[64] The judgment offers significant insights regarding the processing of passenger data and the subsequent automated decisions under the PNR Directive.[65] These insights are also pertinent to the overall approach of the thesis towards ADM.[66]

Regarding the applicability of the transparency model and other findings of the thesis, this part also contains an evaluation of the emerging EU regulatory landscape on transparency and contestation of ADM.

[60] Ch.6. sec. 2.2.1.
[61] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).
[62] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).
[63] Ch.6. sec. 2.2.2 and 2.2.3.
[64] Grand Chamber of the European Court of Justice (ECJ) on 21.06.2022 21 (C-817/19).
[65] Directive (EU) 2016/681 of the European Parliament and of the Council of 27 April 2016 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime.
[66] Ch.6. sec. 2.2.4.

Chapter 2 (*First paper*)

The 'Rule of Law' implications of data-driven decision-making: A techno-regulatory perspective

• Emre Bayamlıoğlu and Ronald E. Leenes, "The "Rule of Law" implications of data-driven decision-making: A techno-regulatory perspective" (2018) *Law, Innovation and Technology*, 10:2, 295 313.

Machines are the concealed wishes of actants which have tamed forces so effectively that they no longer look like forces. [*]

[*] Bruno Latour, *The Pasteurization of France* (Translated by Alan Sheridan, and John Law), Cambridge, Mass.: Harvard University Press, 1993, 204. (Originally published in 1984 as *Les microbes: guerre et paix suivi de imiductions*, A. M. Metailie, Paris)

LAW, INNOVATION AND TECHNOLOGY
2018, VOL. 10, NO. 2, 295–313
https://doi.org/10.1080/17579961.2018.1527475

Routledge
Taylor & Francis Group

The 'rule of law' implications of data-driven decision-making: a techno-regulatory perspective

Emre Bayamlıoğlu and Ronald Leenes

Tilburg Institute for Law, Technology, and Society (TILT), Tilburg Law School, Tilburg University, Tilburg, Netherlands

ABSTRACT

Techno-regulation is a prominent mechanism for regulating human behaviour. One type of techno-regulation concerns automated decision-making with legal effects. While automated decision-making (ADM) systems in the public domain have traditionally been based on conscious design of decisional norms, increasingly, Data Science methodologies are used to devise these norms. This data-driven approach causes frictions with the underlying principle of public-sector decision-making, namely adherence to the rule of law. In this paper we discuss three major challenges data-driven ADM poses to the Rule Law: law as a normative enterprise, law as a causative enterprise and law as a moral enterprise.

ARTICLE HISTORY Received 30 March 2018; Accepted 20 August 2018

KEYWORDS Techno-regulation; automated decision-making; rule of law

1. Introduction

Since the industrialisation, we have witnessed an influx of novel artefacts, objects, and more recently automated systems that come to play a significant role in what we do, how we perceive and interpret the world, how we make our choices, and under what conditions.[1] We have entered an era in which algorithmic systems based on Big Data capitalise economic and institutional power with profound effects on the allocation of resources owing to their capacity to control and manage processes.[2] We see the emergence of 'algorithmic authority' as the legitimate power of 'code' to direct human action and also to impact which information is considered true.

CONTACT Emre Bayamlıoğlu emre.bayamlioglu@uvt.nl

[1]Paul Verbeek, *What Things Do – Philosophical Reflections on Technology, Agency, and Design* (Robert P Crease, tr) (The Pennsylvania State University Press, 2005).

[2]Michael Latzer and others, 'The Economics of Algorithmic Selection on the Internet' (Working Paper, University of Zurich, 2014): http://www.mediachange.ch/media/pdf/publications/Economics_of_algorithmic_selection_WP_.pdf. For more on Big Data and media/information economics, see C Argenton and J Prüfer 'Search Engine Competition with Network Externalities' (2012) 8 *Journal of Competition Law & Economics* 73.

Issues surrounding (big) data analytics and automated decision-making (ADM), such as those touching on privacy and data protection, have been widely studied, but the enabling and restricting role of data-driven solutions as techno-regulatory orders have remained mostly unanalysed.[3] Although studies on techno-regulation frequently analyse and characterise technology for its normativity[4], research theorising the regulatory relevance of Big Data analytics as a normative order in itself is much sparser.[5] As the world of data has become the test bed for social sciences, economic innovation and state administration, the need for research explaining and framing the regulatory dimension of the data-driven practices is ever more critical.

This article contributes to this venture. It departs from the premise that data-driven ADM processes, governed by complex algorithms, are either embodiments of existing normative orders, or they themselves enact *ad hoc* regulatory orders with or without a legal basis. In terms of regulatory constraints and capacities, data-driven ADM systems go much beyond existing legal decision-making based on codified legal norms. Although both types of systems (data-driven *versus* code-driven as Mireille Hildebrandt calls them[6]) regulate human behaviour, their assessment from a rule of law perspective is different. In fact, data-driven ADM systems undermine the rule of law and hence, developers, lawyers and subjects of decisions by these systems should pay attention.

The paper is organised as follows. First, in Section 2, we revisit techno-regulation as a mechanism to regulate human behaviour and describe how conscious implementation of norms is being augmented or replaced by norms derived from data analytics. Next, in Section 3, we discuss some shortcomings and effects of this turn towards data-driven ADM. Section 4 addresses the challenges that these shortcomings cause for the rule of law as the backbone of legal decision-making. Section 5 concludes the paper with some reflections and a call for action.

2. A new horizon of techno-regulation: big data automated decision-making

Left to itself, cyberspace will become a perfect tool of control.[7]

[3]A recent remarkable exception is Timothy D Robinson, 'A Normative Evaluation of Algorithmic Law' (2017) 23 *Auckland University Law Review* 293.

[4]See Lawrence Lessig's *Code and Other Laws of Cyberspace* (Basic Books, 1999) and the descendant literature; WN Houkes, 'Rules, Plans and the Normativity of Technological Knowledge' in MJ de Vries and others (eds), *Norms in Technology* (Springer Science+Business Media Dordrecht, 2013).

[5]M Hildebrandt, 'Law at a Crossroads: Losing the Thread or Regaining Control? The Collapse of Distance in Real Time Computing' in Morag Goodwin, Bert-Jaap Koops and Ronald Leenes (eds), *Dimensions of Technology Regulation* (Wolf Legal Publishers, 2010) 165; Mireille Hildebrandt and Bert-Jaap Koops, 'The Challenges of Ambient Law and Legal Protection in the Profiling Era' (2010) 73 MLR 428.

[6]Mireille Hildebrandt, 'Algorithmic Regulation and the Rule of Law' (2018) 376 *Philosophical Transactions of the Royal Society A*, doi:10.1098/rsta.2017.0355.

[7]Lawrence Lessig, *Code and Other Laws of Cyberspace v.2.0* (Basic Books, 2006) 6.

As the world we are living in becomes densely populated with coded objects, it seems almost 'axiomatic' that the environment and artefacts possess certain governance mechanisms which steer behaviour both at the individual and institutional level – by facilitating or imposing some forms of use and conduct, while inhibiting others.[8] Some have even claimed that technology is law.[9] In a literal sense this is not correct, because law, or legal regulation, is enacted by the legislator and the public bodies that act on the basis of competences attributed by the constitution or the legislator itself.[10] When regulation is taken in the broadest sense to mean intentional influencing of behaviour to produce certain identified outcomes – brought into effect either by code, laws, self-regulation, or by various private schemes[11] – it becomes clear that, from a functional standpoint, both technology and Law may act as regulatory mechanisms which seek to subject human conduct to the governance of certain rules.[12]

Regulation so defined is conceptually closer to the usage in biology, systems theory and cybernetics – encompassing almost any control apparatus or procedure.[13] In fact, Murray and Scott bring control theory into the analysis of regulation. Not only should we be aware of the different modalities of regulation, elaborating on Lessig's famous four (law, norms, market, code), but also that there are three elements necessary to generate a control system: standard-setting, information gathering, and behaviour modification.[14]

[8]This, in fact, is not a new realisation. Jeremy Bentham already in 1787 wrote 'Morals reformed … the gordian knot of the poor-law not cut, but untied – all by a simple idea in Architecture', Panopticon in: Mairan Booi (ed), *The Panoptic Writings* (London: Verso, 2011) 29–95.

[9]Langdon Winner, *Of Autonomous Technology: Technics-Out-of-Control as a Theme in Political Thought* (The MIT Press, 1977) 323–25. Also see Langdon Winner, 'Do Artifacts Have Politics?' (1980) 109(1) *Daedalus* 121. Lessig (n 7) 6.

[10]This is what the rule of law is about. About the 'legal' interpretation of code, see for instance L Asscher, '"Code" as Law. Using Fuller to Assess Code Rules' in Egbert Dommering and Lodewijk Asscher (eds), *Coding Regulation – Essays on the Normative Role of Information Technology* (TMC Asser, 2006) 61–90.

[11]Julia Black, 'Critical Reflections on Regulation' (2002) 27 *Australian Journal of Legal Philosophy* 1; Ronald Leenes, 'Framing techno-regulation: an exploration of state and non-state regulation by technology' (2011) 5 *Legisprudence* 147; Ian Brown and Chris Marsden, *Regulating Code. Good Governance and Better Regulation in the Information Age* (Cambridge, MA, London: MIT Press, 2013). For more on 'regulation', see Roger Brownsword and Morag Goodwin, *Law and the Technologies of the Twenty-First Century. Text and Materials* (Cambridge University Press, 2012); J Kooiman (ed), *Modern Governance* (London: Sage, 1993); C Hood, *The Tools of Government* (London: Macmillan, 1983). For the range and scope of different definitions of regulation, see Lyria Bennett Moses 'How to Think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target' (2013) 5 *Law, Innovation and Technology* 1.

[12]Hans Kelsen, 'The Law as a Specific Social Technique' (1941–42) 9 *University of Chicago Law Review* 75, 79.

[13]Christopher Hood and others, *The Government of Risk: Understanding Risk Regulation Regimes* (Oxford University Press, 2001).

[14]A Murray and C Scott, 'Controlling the New Media: Hybrid Responses to New Forms of Power', (2002) 65 MLR 491, 500. Also, see Andrew D Murray, 'Conceptualising the Post-Regulatory (Cyber)state' in Roger Brownsword and Karen Yeung (eds), *Regulating Technologies* (Hart, 2008) 292.

Techno-regulation refers to the intentional influencing of individuals' behaviour by embedding norms into technological systems and devices.[15] Depending on the context, such regulatory models may interchangeably be referred to as: 'regulation by technology', 'technological normativity', 'regulative software', 'law as design', 'design-based regulation' or 'algorithmic regulation'. Techno-regulatory settings may focus on products/services, places or persons covering a complex plethora of practices and designs. Today, we commonly experience these in driving controls in cars, internet filtering, Digital Rights Management systems, speed bumps, personalised information services, etc. Increasingly, techno-regulation also finds its way in systems that take decisions about individuals and create legal effects.

Vast amounts of raw data compiled from various sources (eg communication networks, the energy grid, and transportation and financial systems) in every realm of life are put to use in order to obtain actionable information for the purposes of detecting of fraudulent transactions, calculation of creditworthiness, organising of Facebook newsfeed and so on. Apparently, our society is heavily dependent on databases and analytic tools to carry out processes of various kinds and scale. Although data-driven practices have long made their way into our lives through statistics and actuarial methods (since at least the nineteenth century[16]), what is happening now is the intense and exponential expansion of these practices by means of the methodologies conceptualised under the term 'big data analytics'. Computational operations for abstraction, correlation, classification, pattern recognition, profiling, modelling, and visualisation are used in a functional way to extract signals from noise in large bodies of data so that those signals can serve as data representations for classifying persons, events or processes.[17] These representations (and profiles) are then used to control processes and make decisions.[18]

[15]Van den Berg and Leenes emphasize and draw attention to other less 'legal' forms of influencing behaviour such as persuasion, or nudging. See Bibi van den Berg and Ronald Leenes, 'Abort, retry, fail: scoping techno-regulation and other techno-effects', in Mireille Hildebrandt and Jaenne Gakeer (eds), *Human Law and Computer Law: Comparative Perspectives* (Springer, 2012). They argue, at 74, that 'persuasion, nudging and affording are more subtle, yet clearly intentional, forms of affecting human behaviour, through the use of technologies, which are overlooked in the current debate on techno-regulation'.

[16]See for instance, Alain Desrosieres, *The Politics of Large Numbers: A History of Statistical Reasoning* (Camille Naish, tr). Originally published as La politique desgrands nombres: Histoire de la raison statistique (Editions La Decouverte, 1993).

[16]See for instance, Alain Desrosieres, *The Politics of Large Numbers: A History of Statistical Reasoning* (Camille Naish, tr). Originally published as La politique desgrands nombres: Histoire de la raison statistique (Editions La Decouverte, 1993).

[17]Jerry Kaplan, *Humans Need Not Apply, A Guide to Wealth and Work in the Age of Artificial Intelligence* (Yale University Press, 2016) 25.

[18]KEC Levy, 'Relational Big Data' (2013) 66 *Stanford Law Review Online* 73, n.3; Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt Publishing Company, 2013).

Data analytics has become a method of empirical inquiry, performed on informational sources to extract new insights out of raw data, supplementing or even substituting the conscious design of rules to control processes and decisions; thus moving from causation as the link between input and output to correlation.[19] Conceptualising big data as a methodology – rather than as a computational source/tool/instrument defined with reference to size and speed – provides a framework which enables the analysis of the regulatory aspects of data-driven methodologies, and the ensuing rule of law implications that will be elaborated in the following parts of this paper.

Regulation, standard-setting, monitoring and behaviour modification by means of computational algorithms is nothing new.[20] Governmental bodies have used algorithms in decisional processes since the dawn of the computers. Levying taxes, and more generally, the social welfare state, would not be possible without these automated decision systems.[21] The way legislation is transformed into executable code is what is new.

One classical approach has been to represent state-of-the-art domain knowledge in production rules (if–then rules), and then have an inference engine reason on these to give expert-like advice or make decisions.[22] In many of these legal knowledge based systems (LKBS) – a relatively successful type of rule-based application – developers represented 'the law' in executable form. This allowed the systems to make correct legal decisions and be able to explain or legally justify their reasoning process together with the conclusions they reached.[23] The developers of such systems aimed at faithfully representing the authoritative legal source in the domain of application as well as the anticipated kinds of cases relevant to the domain (and rule-based representations of existing case law).

This approach, however, never really caught on substantially. Quite apart from requiring significant effort to represent legal rules, which affected the adoption of this methodology of building (A)DM systems, there are also

[19]Michael Mattioli, 'Disclosing Big Data' (2014) 99 *Minnesota Law Review* 538.
[20]Cf. Hildebrandt (n 6) 2.
[20]Cf. Hildebrandt (n 6) 2.
[21]While we focus on automated decision systems, in the end the same reasoning applies to advice giving systems. See Hildebrandt (n 6); Jason Millar and Ian Kerr, 'Delegation, Relinquishment and Responsibility: The Prospect of Expert Robots', in Ryan Calo, Michael Froomkin and Ian Kerr (eds), *Robot Law* (Edward Elgar, 2016) 102–28, on the inevitability of relinquishing control to machines.
[22]These types of systems have been in operation since the 1970s. See for instance, EA Feigenbaum, 'The Art of Artificial Intelligence: I Themes and Case Studies of Knowledge Engineering. Technical Report' (UMI Order Number: CS-TR-77–621, Stanford University, 1977); Andrew Stranieri and John Zeleznikow, *Knowledge discovery from legal databases* (Springer, 2010).
[23]'Not necessarily through mimicking the actual reasoning process, but by, for instance, implementing the underlying (complex) legal rules and executing those'. Trevor Bench-Capon, 'Exploiting isomorphism: development of a KBS to support British coal insurance claims', Proceedings of the 3rd International Conference on Artificial Intelligence and Law, New York, 1991, 62–68; Jörgen Svensson 'Legal expert systems in general assistance: from fearing computers to fearing accountants' (2002) 7 *Journal of Information Polity* 143. Also, on the failures of LKBS, see P Leith, 'The rise and fall of the legal expert system' (2010) 1 *European Journal of Law and Technology* (Issue 1).

fundamental problems due to the intentional open-texturedness and vagueness of the human language through which the law is expressed. Moreover, the application of legal rules is highly context dependent, meaning that the fringes of what such a regulatory mode appropriately handles are easily reached.[24] The LKBS approach is limited due to the difficulty of dealing with fundamental characteristics of legal norms (open-texture, vagueness)[25] and its inherent difficulty to cope with the dynamics of the domain it purports to govern.[26] A further complication is that many, if not all, domains in which legal decisions are taken are characterised by a combination of 'positive' law and 'case' law.[27] The rule-based LKBS approach, due to its rule based nature, has difficulty in coping with dynamic case law.

Owing to the advances in the fields of data analytics, semantic web and Natural Language Processing (NLP), data-driven ADM systems are now beginning to assign meaning to vague terms, and 'interpret' normative standards, and principles to 'manage' the uncertainties of the human language by deriving knowledge from a large legal corpus including the case law.[28] Modern techniques could potentially overcome the static (and limited) nature of the classical rule-based LKBS because of their adaptive capacities and affordances. Rule-based (code-driven) systems, by incorporating data analytics capabilities, may mitigate the rigidness of pre-set architectures – implementing norms by way of incorporation of new knowledge through (machine) learning and feedback mechanisms and thus become data-driven.

Since techno-regulation is defined as the effectuation of norms through technical means at various levels such as rule-making, implementation, monitoring and enforcement in a normative system, the intrinsic regulatory capacity of data-driven ADM is evident. We see the regulative force of data analytics in almost every context where operation or conduct of certain activity is, either fully or partially, automated or controlled by algorithmic decision-making systems.[29] The predictive and the pre-emptive nature of

[24]Lyria Bennett Moses and Janet Chan, 'Using Big Data for Legal and Law Enforcement Decisions' (2014) 37 UNSWLJ 643, 657.

[24]Lyria Bennett Moses and Janet Chan, 'Using Big Data for Legal and Law Enforcement Decisions' (2014) 37 UNSWLJ 643, 657.

[25]Abdul Paliwala, 'Rediscovering artificial intelligence and law: an inadequate jurisprudence?' (2016) 30 *International Review of Law, Computers & Technology* 107; Philip Leith 'The Rise and Fall of the Legal Expert System' in Abdul Paliwala (ed), *A History of Legal Informatics* (Prensas de la Universidad de Zaragoza, 2010) 179–203.

[26]See Ronald Leenes, 'Hercules of Karneades: Hard cases in recht en rechtsinformatica' (Universiteit Twente, 1999) (in Dutch).

[27]We put positive law and case law in quotes to signify that both sources are not limited to material produced by the legislative and judicial branches of government. Rather, we mean authoritative rules that are adjudicated (or enforced) by some agency that has the authority to do so.

[28]See Kevin Ashley, *Artificial Intelligence and Legal Analytics – New Tools for Law Practice in the Digital Age* (Cambridge University Press, 2017).

[29]Karoline Krenn, 'Markets and Classifications – Constructing Market Orders in the Digital Age: An Introduction' (2017) 42(1) *Historical Social Research* 7, 15: http://dx.doi.org/10.12759/hsr.42.2017.1.7-22.

data analytics amplify both the direct and indirect regulative impact of the ICTs.[30]

The resulting systems could take the form of a combination of classical, including handcrafted, rule-based representations augmented with knowledge derived by Machine Learning (ML). In any case, these systems are capable of dynamically adapting to their environment owing to the complex data-driven knowledge bases that are not directly intelligible.

3. Data-driven ADM concerns, challenges and potential harms

Data-driven ADM processes, governed by algorithms of varying degrees of complexity are either the embodiment of existing normative orders, or they themselves enact *ad hoc* regulatory orders with or without legal basis such as the case of online advertising where algorithms decide who is worthy of receiving a discount, or the call service using sentiment analysis to decide which of the callers is more tolerant to be kept waiting.[31] Although such trivial practices may seem irrelevant from the legal perspective, a second thought reveals several repercussions with regard to consumer rights and human dignity in general.

It should also be borne in mind that there are secondary effects. ADM does not necessarily involve decisions directly about the individuals. For instance, a simple ML application to recognise congestion on visual data (eg from a traffic surveillance camera) may give rise to biased decisions with regard to traffic flow, depending on the data and the way of processing. One other dimension is that nothing comes for free, that is, the efficiency gains or other benefits to be derived from data analysis also have trade-off effects in other domains or for other individuals. Cutting costs through data analysis could mean certain economic and material diversions, and a shift of interests among employees, students, citizens or consumers. For instance, reducing the cost of handling customer complaints through a techno-regulatory application (eg automated classification and diverting of complaints to the relevant departments) may give rise to a significant change in a company's way of communicating with the public. Moreover, such systems – though not necessarily intentionally – run the risk of favouring certain type of complainants against others without any just cause. Or, a bank which decides to use predictive analytics to prevent customer churn can act pre-emptively such as to offer advantageous services to the customer who is regarded to be more likely to move to another bank. This may seem to be a discriminatory result in that many of us would not consider risk of churn as a legitimate basis on the side of the bank to differentiate between the service receivers.

[30]Ian Kerr and Jessica Earle, 'Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy' (2013) 66 *Stanford Law Review Online* 65.
[31]Luke Dormehl, *The Formula: How Algorithms Solve All Our Problems and Create More* (WH Allen, 2015).

ADM, when coupled with data analytics, acquires the necessary adaptive capability to diffuse into more general domains controlling and regulating real-life events that are of relevance to law and to the legal system.

The emergence of 'algorithmic regulation' legitimises the power of the 'code' to direct human action. But with this, the risks of epistemological flaws and biases inherent to machine learning enter the scene. These may raise concerns as to fairness/non-discrimination, privacy/invasiveness, and the notions of the 'autonomous self' and dignity.

Machine learning is a problem-solving approach which implements statistical learning theory as a framework of computational strategies for discovering 'truth' in empirical questions. Data mining employs quantitative and inductive methods (equations and algorithms), along with statistical testing to process data resources with a view to identifying reliable patterns, trends, and associations among variables that describe and/or anticipate a particular process or event. What can be derived from the data is determined by what is in the data, what the system designers label as the relevant factors to be analysed, and the adopted methodologies. For instance if the training dataset for predicting court decisions consists of case law, a relevant question is *which* cases are incorporated in the corpus. Does it feature all decided cases or only those that were published (and hence selected by an editorial board)? What material related to the case is taken into account? All files, or only the judgment? In the latter case, one has to be aware that the facts may be formulated to align with the conclusion reached in the case.[32]

Data are not capable of verifying the assumptions and the perspective underlying a certain inference of causation. So, letting data speak for itself thus is problematic in many ways. Algorithms in machine learning are not immune from the general shortcomings of the causal inference in large data sets. Data mining reveals correlation, not causality, which could be spurious, and this brings in the question of the ethical justifiability of acting upon them.[33] In order to establish a causal link, patterns need models with an encompassing narrative since 'it is one thing to establish significant correlations, and still another to make the leap from correlations to causal attributes'.[34] As an inductive method – progressing from particular cases

[32]N Aletras and others, 'Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective' (2016) *PeerJ Computer Science* 2:e93 https://doi.org/10.7717/peerj-cs.93, cited in Hildebrandt (n 6).

[33]'Episcopalian dog owners who drive more than forty miles to work and recently moved to the suburbs may have an extraordinarily high rate of bladder cancer, but so what? The correlation is probably spurious. Nothing about dog ownership, being Episcopalian, or recently moving to the suburbs would seem to cause bladder cancer. The challenge is to sort through all of the correlations and decide which have a causal basis', Scott E Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton University Press, 2008) 85.

[34]David Bollier, 'The Promise and Peril of Big Data' (Aspen Institute) 2010, 16.

(sample data) – machine learning accumulates a set of discovered dependencies, correlations or relationships that are referred to as 'model'. Although a model in the abstract may be robust and consistent, it may nevertheless be favouring certain values, persons, or processes – bringing us to a domain which is more political, rather than being scientific.[35]

A well cited example of legal analytics that indirectly shows bias and epistemological flaws is the study performed by Roger Guimerà and Marta Sales-Pardo, who devised a model to predict a justice's vote (in the US Supreme Court) based on the other justices' votes in the same case.[36] The model predicts votes more accurately (83%) than human experts. However, the model does not take into account the content of the case, but only 'metadata'. In another often cited study, researchers built a model to predict the outcomes of the 2002 Term. Again, the system outperformed (with 75% accuracy) expert predictions. And again, no information about the case or applicable law was incorporated in the model. Instead, features like the name of the judge, the term, the issue, the court of origin and whether oral arguments were heard were used.[37] Both studies illustrate how the normative force of the law – that was present in code-driven systems – becomes replaced by the patterns in a (historic) dataset that may have nothing to do with legal norms.

4. The rule of law implications

Technology is never neutral,[38] yet in the eyes of many, technology and politics are separated in that politics is supposedly based on values, while technology thrives on scientific knowledge and objective facts.[39] It propagates an interpretation of regulation from an external perspective, which focuses on behavioural modification (by any means), while neglecting the internal perspective that deals with checks and balances of the rule of law. An apparent result of such dualism is the lack of democratic control over much techno-regulation. Whereas law is created in the public domain, techno-regulation (even when adopted by 'the state') often is not.[40] Yet, techno-regulation

[35]Lucas Introna, and Niall Hayes 'On Sociomaterial Imbrications: What plagiarism detection systems reveal and why it matters' (2011) 21 *Information and Organisation* 107, 108.
[36]R Guimerà and M Sales-Pardo 'Justice Blocks and Predictability of U.S. Supreme Court Votes' (2011) PLoS ONE 6(11): e27188. https://doi.org/10.1371/journal.pone.0027188.
[37]Theodore W Ruger and others, 'The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking' (2004) 104 *Columbia Law Review* 1150.
[38]Mireille Hildebrandt, 'A Vision of Ambient Law' in Roger Brownsword and Karen Yeung (eds), *Regulating Technologies* (Hart, 2008) 175–92; Winner, 'Do Artifacts Have Politics?' (n 9).
[39]A Feenberg, 'Critical Theory of Technology' in JKB Olsen and others (eds), *A Companion to the Philosophy of Technology* (Blackwell Publishing, 2009) 149. Also see M Bunge, *Evaluating Philosophies* (Science +Business Media Dordrecht, 2012) 5.
[40]Leenes, 'Framing Techno-Regulation' (n 11) 147–48.

should be situated in a wider framework encapsulating the mutual entanglements between culture, politics and technology. As Don Ihde has put: 'technological form of life is part and parcel of culture, just as culture in the human sense inevitably implies technologies'.[41] Or, as Andrew Feenberg writes 'Technology should be brought into the public sphere where it increasingly belongs'.[42]

Every legal system has a claim to legitimacy in the sense that the source of authority relies on a moral right to rule.[43] In modern democratic systems, the principle of the rule of law, as an essential pillar of this moral dimension, requires that rules are publicly declared with prospective application, and possess the characteristics of generality, equality, and certainty.[44] As the protection of rights, prevention of arbitrariness and holding the state responsible for unlawful acts are only possible in an intelligible, reliable and predictable order, universality and relatively constant application over time in a prospective and non-contradictory way may be regarded as the main constituents of the notion of rule of law.[45] Rights are of little use if their limits and proper scope are not in advance known by citizens.

An important procedural dimension of the rule of law, which is of particular concern from the ADM perspective, is the effective capability to contest decisions.[46] This primarily requires that one must be aware of the existence of an ADM process, and also foresee and understand the consequences.[47] Law's capacity to allow subjects to contest judicial and administrative decisions, including the validity of the rule itself, provides a meta-level procedural safeguard in that 'the addressees and the "addressants" of legal norms coincide' – a form of self-regulation where the law maker is bound by the rules of its own creation.[48]

Against this backdrop, we conceptualise three potential harms of data-driven techno-regulation which undermine the rule of law as a procedural safeguard to discern, foresee, understand and contest decisions – namely (i) the collapse of the normative enterprise (ii) the replacing of a causative basis with correlative

[41] Don Ihde, *Technology and the Lifeworld. From Garden to Earth* (Indiana University Press, 1993) 20.
[42] Feenberg (n 39).
[43] 'Or as Thomas Hobbes might have put it, how is authority now authorized?' Zygmunt Bauman and others, 'After Snowden: Rethinking the Impact of Surveillance' (2014) 8 *International Political Sociology* 121.
[44] Brian Tamanaha, *On the Rule of Law* (Cambridge University Press, 2004).
[45] Jeremy Waldron, 'The rule of law in contemporary liberal theory' (1989) 2 Ratio Juris 84; Hans-Wolfgang Arndt, 'Das Rechtsstaatsprinzip' (1987) 27 JuS L41–L44.
[46] Speaking of natural overlaps between the substantive and procedural aspects of the rule of law, Waldron mentions that a hearing by an impartial tribunal acting on the basis of the evidence and arguments presented, a right to hear reasons from the tribunal when it reaches its decision, and some right of appeal to a higher tribunal as procedural characteristics are equally indispensable. Jeremy Waldron, 'The Rule of Law and the Importance of Procedure', in James E Fleming (ed), *Getting to the Rule of Law* (New York University Press, 2011) 7.
[47] M Hildebrandt, 'Profile transparency by design? Re-enabling double contingency' in M Hildebrandt and K de Vries (eds), *Privacy, Due Process and the Computational Turn* (Routledge, 2013).
[48] Mireille Hildebrandt, *Smart Technologies and the End(s) of Law* (Edward Elgar, 2015) 10.

calculations, and (iii) the erosion of moral enterprise.[49] The informational asymmetries, flawed epistemology of data-driven inferences together with the bias inherent in machine learning of such regulation bring about the concern that the 'rule of law' might be exchanged for the 'rule of technology' – accompanied by *Kafkaesque, Huxleyan* and *Orwellian* discourses of dystopia.[50]

4.1. Challenge to law as a normative enterprise

Rules, principles, standards and in general 'norms' provide uniformity, predictability, and social coordination for they inform individuals about their way of conduct, and explain the legal course of events in situations addressed by the Law. Law, hence, is a normative enterprise where the legislator consciously creates legal effects (institutional facts) that obtain when certain conditions are met.[51]

Any regulator will weigh various interests and decide what the norm should be in a particular constellation of facts. The norm is usually written down allowing the regulatees to take note of it and act accordingly. Regulatees are supposed to adhere to the norms and if they transgress the norm, face the consequences. However, normativity does not stop here, otherwise enforcing the norms through technology would potentially fully realise the ideal sketched by the law. Statutory norms represent the solidification of a political debate at a particular moment, taking into account only the foreseeable facts, interests and effects. Changing knowledge, opinions, interests etc, may require reopening the debate, and hence contestation of norms is an essential mechanism so that law and society can mutually evolve. Courts will decide how to cope with new arguments and new situations, and how to ensure that their verdict is enforceable and comprises law.

As explained above, there is some implicit normativity in every decision. Any decision-making system has a normative basis which may be seen as a totality of the decisional criteria, assumptions, and legitimations embedded in the system, specifying its behaviour.[52] However, techno-regulatory settings based on data-driven correlations and inferences pose a challenge to law as a normative enterprise in that there are no clear enacted norms in the

[49]This trilogy has been briefly visited in Ugo Pagallo and others, 'New technologies and law: global insights on the legal impacts of technology, law as meta-technology and techno regulation' New-Technologies-and-Law-Research-Group-Paper, 4th LSGL Academic Conference, Mexico 2017.

[50]Roger Brownsword, 'So What Does the World Need Now? Reflections on Regulating Technologies' in R Brownsword and K Yeung (eds), *Regulating Technologies* (Hart, 2008) 23–48. For more on the implications of ML that may disrupt the concept of the rule of law, see Mireille Hildebrandt, 'Law As Computation in the Era of Artificial Legal Intelligence. Speaking Law to the Power of Statistics' *University of Toronto Law Journal* Volume 68 Issue supplement 1, January 2018, 12-35). https://ssrn.com/abstract=2983045.

[51]Brian Z Tamanaha, *A Realistic Theory of Law* (Cambridge University Press, 2017) 121. Also, see Dick W. P. Ruiter, *Institutional Legal Facts: Legal Powers and their Effects* (Springer-Science+Business Media, 1993) 205-207.

[52]MJ de Vries, SO Hansson, and AWM Meijers (eds), *Norms in Technology* (Springer Netherlands, 2013).

conventional sense anymore to provide a mapping between the facts and the legal effects.[53]

In data-driven ADM, decision rules are (partially) *dynamic*. The norms imposed by these systems are not stable, but rather they are the objects of persistent and on-going reconfiguration.[54] The decisional rule itself emerges (autonomously) from the (dynamic) data used for training the system.[55] What is regarded to be the 'norm' is no longer predetermined, but constantly adjusted and opaque (normative opaqueness).[56] As inferential statistics and/or machine learning techniques produce probable yet uncertain knowledge, when statistics instead of reason *de facto* enter into the realm of norm setting, law loses its normative basis – at least to the extent that we associate normativity with human action.

A further type of normative opaqueness is due to the difficulties in discerning the *intention of the rule-maker*. In a data driven setting, the programmer sets the boundaries for learning, but as we have seen extraneous factors may find their way into the decisional rules. The normative impact of the ADM therefore is not solely determined by (legislative) intent. The affected individual cannot discern which part of the normativity (as could be inferred from the output) is intentional and which part is merely spin-off in the form unforeseen or secondary effects. Accordingly, the outcome in a data-driven setting may not be regarded as fully reflecting the intent of the competent body to enact rules.

Added to this is the *computational complexity* of data-driven systems.[57] Algorithms are unintelligible in the sense that the recipient of the output (eg a classification decision) rarely has any concrete idea of how or why a particular classification has been made (even if it is clear what the input was). The self-adjusting and adaptive capacity of data-driven systems renders them

[53] 'As well, the specified variables could be the result of still other forces to which we should pay attention: a statistical model might gain accuracy by including the race, sex, age, and income of the parties, lawyers, and judges participating in a case without revealing precisely why or how these attributes influence decision-making. Useful variables will not necessarily map out decision dynamics'. Adam Samaha, 'Judicial Transparency in an Age of Prediction' (University of Chicago Public Law & Legal Theory Working Paper No. 216, 2008) 9.

[54] See Brent Daniel Mittelstadt and others, 'The ethics of algorithms: Mapping the debate' (2016) 3 *Big Data & Society* (https://doi.org/10.1177/2053951716679679).

[55] Massimo Buscema and William J Tastle (eds), *Intelligent Data Mining in Law Enforcement Analytics – New Neural Networks Applied to Real Problems* (Springer Netherlands, 2013) 14.

[55] Massimo Buscema and William J Tastle (eds), *Intelligent Data Mining in Law Enforcement Analytics – New Neural Networks Applied to Real Problems* (Springer Netherlands, 2013) 14.

[56] 'In contrast to human-made rules, these rules for decisionmaking are induced from historical examples – they are, quite literally, rules learned by example. Joshua A Kroll and others, 'Accountable Algorithms' (2017) 165 *University of Pennsylvania Law Review* 633, 679. Also see Matthias Leese, 'The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union' (2014) 45(5) *Security Dialogue* 501.

[57] Anton Vedder and Laurens Naudts, 'Accountability for the Use of Algorithms in a Big Data Environment' (2017) 31 *International Review of Law, Computer & Technology* 206.

intractable and unintelligible to human cognition.[58] Opacity in machine learning algorithms is a product of the high-dimensionality of data, complex code and constantly reconfigured logic of the decision-making.

4.2. Challenge to law as a causative enterprise

Legal regulation is normative. Legal effects are not a matter of correlation between certain facts and effects, but of (legal) causation, or rather the law creates (constitutes) legal effects. The standard-setter determines which conditions lead to which legal effects. Data-driven ADM systems interfere with this mechanism due to their reliance on correlation.

Data analytics employ quantitative methods and statistical testing to process data resources to identify reliable patterns, trends, and associations among variables that describe and/or anticipate a particular process.[59] As a novel method of empirical inquiry, instead of starting with a question, Big Data reverses this process by first running the algorithms to look for patterns, and then retrospectively constructing hypotheses.[60] The seeming strength and comprehensiveness of this methodology relies on the magnitude of the datasets providing an oligoptic[61] view of full resolution – the belief that 'with enough data, the numbers speak for themselves'.[62]

There are some evident restrictions and limitations of the methodology of extracting knowledge out of patterns and correlations identified in large datasets. First, in large enough datasets, even if data is selected arbitrarily, certain patterns will occur when analysis extends long enough. With so many possible dimensions, it becomes incredibly likely that some constructed type correlates with the outcome.[63]

Some correlations are straightforward; almost axiomatic easy observations – for example, demand for flu medicine increases in winter, and more traffic accidents take place during rain. And some may be more subtle and sinister

[58] Jenna Burrell, 'How the machine "thinks": Understanding opacity in machine learning algorithms' (2016) *Big Data & Society*, 1–12; Antoinette Rouvroy, 'The end(s) of critique: data-behaviourism vs. due-process'; Valeria Ferraris and others Working Paper 'Defining Profiling' (2013) https://www.academia.edu/5398935/Defininig_Profiling; Ronald Leenes and Paul de Hert (eds), *Reforming European Data Protection Law* (Springer Netherlands, 2015); Nicholas Diakopoulos 'Algorithmic Accountability: Reporting On The Investigation of Black Boxes' (Columbia University, 2014): https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2; Jatinder Singh, Ian Walden, Jon Crowcroft, and Jean Bacon, 'Responsibility & Machine Learning: Part of a Process', (October 27, 2016): https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2860048.

[59] Stephan Kudyba *Big Data, Mining, and Analytics* (CRC Press, 2014) 29.

[60] Mattioli (n 19); Chris Anderson, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', *Wired* (23 June 2008), http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory.

[61] Rob Kitchin, 'Big Data, new epistemologies and paradigm shifts', *Big Data & Society*, April–June 2014, 1–12, 4.

[62] Anderson (n 60).

[63] 'Note that it is exactly the size of the data that allows our result: the more data, the more arbitrary, meaningless and useless (for future action) correlations will be found in them'. Cristian S Calude and Giuseppe Longo, 'The Deluge of Spurious Correlations in Big Data' (2017) 22 *Foundations of Science* 595. Also, see Scott E Page, *The Difference* (Princeton University Press, 2007) 85

like overweight persons make more spelling mistakes, while some are simply valuable such as the knowledge that a US citizen is more likely to register to vote after being informed that a close friend has registered. However, a correlation does not necessarily amount to causation[64] – for it does not inform us about the nature of the discovered relation. The correlation between independent and dependent variables in the analysis may be spurious. There may not be a causal relation between diapers and beer, though it may be equally plausible that people buying diapers have kids and therefore they consume beer at home, rather than going out with friends. In such cases, although the supposed cause and effect are related, in fact they may be both dependent on a third factor.

The meaning constructed through repeated observations over time and/or space does not necessarily explain but may undeniably rationalise what otherwise would be regarded as coincidental or unpredictable.[65] The basic premise behind data analytics is that the observation of correlations along the chosen parameters would extend into future events. However, a correlation may be a weak epistemological basis for prediction and thus, the so-called 'truth' offered by Big Data may turn out to be nothing more than a discursive self-intoxication.[66]

Without doubt, certain correlations are useful observations for their practical relevance. However, as the data itself is not capable of justifying the assumptions and the perspective underlying a certain inference , correlations have no causative explanatory link unless narrated through a theory and implemented as a model based on that theory. Even though patterns are detected by algorithms, the input (data), algorithms to be used, and many other design choices make data analytics a model-building exercise.Therefore correlations are not 'just discovered', but also manufactured. This unfolds the further epistemological problem that causality in data-driven practices is a question of model-building which is itself a value-laden theorisation.[67] Thus, every predictive model inevitably discards certain part of the information about the world around us, and by doing so, it enables us to reach a digitised representation of the problem space which can be manipulated by means of algorithms.[68] In order to asses causal value, we need to know the range of alternatives from which a certain interpretation is derived, together with the principles and factors which generate that range of options.

[64]Mayer-Schönberger and Cukier (n 18), ch.1.
[65]A Jacobs, 'The Pathologies of Big Data' (2009) 52 *Communications of the ACM* 36.
[66]Grégoire Chamayou, *A Theory of the Drone* (The New Press, 2015).
[67]Stavros Ioannidis and Stathis Psillos 'Mechanisms, Counterfactuals, and Laws' in Stuart Glennan and Phyllis Illari (eds), *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (Routledge, 2018). Also see Loise Amoore, *The Politics of Possibility* (Duke University Press Books, 2013) 44.
[68]David M Berry, *The Philosophy of Software – Code and Mediation in the Digital Age* (Palgrave Macmillan, 2011).

An epistemology establishing causation between a multitude of data points through aggregation and recursive data analysis – insights of which may not be understood through direct human cognition – signifies the demise of law as a causative enterprise. Such a break of the causation chain is also a serious blow to human autonomy because individuals could no longer contest the result through rational argumentation. The collapse of the causative link may also be seen as a big leap towards dehumanisation of the social, economic, and political texture of our lives.

4.3. Demise of law as a moral enterprise

Data-driven models implementing rules or legal frameworks impair the rule of law by undermining the moral basis of the legal system on many fronts. Frist, the arguments within this context primarily relate to the notions of human autonomy and dignity as the higher principles of European legal and political order since the Enlightenment. Where technology is used to steer human conduct with a view to ensure compliance or for the implementation of certain norms, not only the normative character of law suffers from erosion, but also *human autonomy* and the moral grounds that the very norms are predicated upon. Especially where an *ex-ante* regulatory approach is taken – leaving no room for breach, or choice as to the way of compliance – our thinking of law departs from 'should/should not' to 'can/cannot', meaning that what is not legal cannot be done either.[69] Hence, techno-regulation can take away the freedom to deviate from the embedded norm in various ways.[70] Compare, for instance, the tourniquets found at different train and metro systems around the world. In some cases the barrier is man-high, in others one can easily climb/jump over them. In the first case, transgressing the norm is impossible, in the second the choice between morality and deviance is present.[71] The difference may seem trivial, but taking away the personal choice by rendering certain behaviour impossible may lead to weakening of self-controls and may have a de-moralising effect.[72]

Such erosion of human autonomy is aggravated in the case of data-driven DM models where the norms are not stable, but rather subject to persistent and on-going change and reconfiguration – making a moral anchoring less possible. This malleable and 'fluid' nature of data-driven systems make them particularly attractive as a regulatory tool, but very unattractive from

[69]While ex-post methodologies discourage non-compliance or improve the chances of detection, without eliminating individual choice, the ex-ante approach overrides the individual as an intentional agent and automatically imposes the desired state or pre-empts certain behavior. See Kerr and Earle (n 30).

[70]Leenes, 'Framing Techno-Regulation' (n 11); K Yeung, 'Can we Employ Design-Based Regulation While Avoiding Brave New World?' (2011) 3 *Law, Innovation and Technology* 1, 2.

[71]K Yeung, 'Towards an Understanding of Regulation by Design', in R Brownsword and K Yeung (eds) (n 50) 98.

[72]DJ Smith, 'Changing Situations and Changing People', in A von Hirsch, D Garland and A Wakefield (eds), *Ethical and Social Perspectives on Situational Crime Prevention* (Hart, 2000).

the perspective of agent morality – eliminating the opportunities to act in a moral way by one's own will and thus undermining the conditions required for a flourishing moral community.[73] As explained above, although data-driven approach may cure the giddiness of rule-based systems to ensure 'efficient' compliance and execution, such positive gains are achieved at the expense of individual autonomy and agent morality. The adaptive and pre-emptive capacity of data-driven systems deprives individuals of the ability to reason with the rules.

Second, the application of Data Science techniques in the legal domain has been described as an important factor that may change how the legal services operate as well as the way the judiciary functions.[74] The core idea here is that data-driven legal analytics trained on data extracted from 'legal sources' such as case law and even doctrinal research will allow the construction of systems that will predict legal consequences with high precision—renderingthe process of adjudication almost idle. Some even believe that a 'legal singularity' is near because the ' … accumulation of massively more data and dramatically improved methods of inference make legal uncertainty obsolete.'[75] Whatever one may think of the feasibility of this, it may be the case that application of data analytics on the existing case law may produce a model that is able to accurately predict the outcome of every case that falls within the boundaries of the training set.[76] Indeed, the performance of systems trained on a set of cases may be good in the sense of accurately predicting the outcome of a case relative to its body of knowledge (the training set).[77] The outcomes of cases not covered by the training set are speculative and it is unknown whether these judgments are 'legally correct'.[78] In other words, the model

[73]R Brownsword, 'Code, Control, and Choice: Why East is East and West is West' (2005) 25 *Legal Studies* 1, 17.

[74]See, for instance, Richard and Daniel Susskind, *The Future of the Professions* (Oxford University Press, 2015); Daniel Martin Katz, 'Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry' (2013) 62 *Emory Law Journal* 909.

[75]Alarie Benjamin, 'The Path of the Law: Toward Legal Singularity' (May 27, 2016). https://ssrn.com/abstract=2767835.

[76]This is a fundamental problem in AI and Law, known as the frame problem. Within the boundaries of the knowledge of the system, its performance may be good, but the system will not be able to handle cases outside these boundaries, nor will it generally be able to detect that a case actually falls outside its frame of knowledge/reference. It operates on a closed world assumption. Law, however, is a dynamic open system, engaging potentially with any case outside the system's perimeters. See Leenes (n 26).

[77]In other words, these models do not really predict, but rather describe a historical data set, see Hildebrandt (n 6) 7.

[78]The system can thus handle 'clear cases' as they are called in legal theory (see Dworkin), not 'hard cases', which can be taken to mean here cases that fall outside the frame of the system, or cases that are made to fall outside the frame by contestation. Nor does it notice a hard case has been presented to it. As a result of contestation, any case, also seemingly clear cases (or cases that are treated as clear by the system), may be turned into hard ones, for which the system may produce the wrong result. Moreover, even a perfect system (the magical algorithm, the point of legal singularity) will have diminishing returns, as the confidence of the system will be impaired by the decreased number of new cases to observe due to decreased need for adjudication. However, if seen from the perspective of cybernetics, this positive feedback may be offset in that the system's loss of reliability in time will result in more

can retrospectively predict the outcome of legal disputes only within a very limited understanding of what the law is about. As this may seem unproblematic and even laudable for helping the under-privileged access legal advice or facilitating the extra judicial settlement of disputes, Hildebrandt and others have rightly pointed out:[79]

> [...] law must be understood as a coherent web of speech acts that inform the consequences of our actions, itself informed by the triple tenets of legal certainty, justice and instrumentality that hold together jurisdiction (the force of law), community (even if between strangers) and instrumentality (the policy objectives of the democratic legislator).

The magical algorithm may render the law fully predictable, but it will still lack the necessary transparency and moral accountability in the sense of being open to scrutiny, and consequently compliant with the rule of law.[80] For being an affront to man's dignity as a responsible agent, replacing adjudication processes with predictable outcomes is a significant impairment to the rule of law for it undermines the moral premises of the legal system.

'Mathematical simulation of legal judgement' should not be mistaken for the judgment itself.[81] Where decisions are not contestable through argumentation, there exists no authority to morally defend and justify the decision. Even if we knew that the analytics provide the best possible solution, and accurately predict the outcome of every possible dispute in advance, we would still need to render such decision intelligible so that it is transparent enough to be contested. Although such magical algorithm appears to relieve us from the burden of arguing cases before the courts, this does not in fact suppress the need for argumentation as a moral justification process. Delivery of an explanation to substantiate any decision is crucial in obtaining the necessary acceptance and endorsement from the individuals who are subject to the system. Adjudication not only provides redress but also has a connotation of morality through explanations that render the outcome normatively acceptable. The idea of predictive judgment, which eliminates the need for adjudicatory process, discards this moral signalling function of law.

5. Conclusion: conflicts to paradoxes

The pervasive employment of data-driven systems is indicative of our current and future dependence on technologies incorporating, articulating and

disputes being taken to court – eventually pushing the system back to perfection with the introduction of fresh data. Accordingly, instead of replacing the judiciary, predictive analytics may be used as a tool to monitor and audit actual court decisions.

[79] Hildebrandt (n 6).

[80] Samaha (n 53).

[81] Hildebrandt (n 6).

amplifying computational and calculative rationalities – linking ends to means in novel and humanly unintelligible ways.

Counting, calculating, accounting and eventually computing – a hectic obsession of modern humans – now has reached the point where we turn blind to almost anything that falls beyond or outside of our measuring capacity.[82] The social complexity we live in dictates a paradigm where knowledge is limited without measurement.[83] This current prevailing understanding of data analytics and technology is rooted in the political philosophy of modern societies which is predicated upon a distinction between *politics* and *science*, according to which, while the former is supposedly based on values, the latter seeks for "objective truth".[84]

The problem with the emerging data-driven epistemology is that the kind of *knowing* it suggests is not always what we aim for or desire if we want to maintain the rule of law, but simply what technology allows us. Or as David Berry put it: 'subtractive methods of understanding reality (*episteme*) produce new knowledges and methods for the control of reality (*techne*)'.[85]

Data-driven processes increasingly re-embody norms within a form of an instrumentalized rationality. Data-driven instrumental reason converts each dilemma, conflict or antagonism, however material and fundamental, into a mere paradox which could be counteracted by the application of logic – substituting interests with the requirements of the technique and the normativity of law with the performativity of the algorithm. Big data constrains the possibilities for political and moral choices by reducing governance to a technical process of adaptation, and law to a process of optimisation – rendering politics a mere question of "better-doing".[86]

If the rule of law is taken as a meta-principle which primarily presupposes an autonomous subject who could effectively reason against the norms and introduce a novel interpretation,[87] the type of law that the data-driven paradigm implements, leaves no room for effective contestation – but only rationalised logical and probabilistic reasoning. This results in an all or nothing approach which hardly complies with the principles of proportionality, subject autonomy, expediency and certainty.[88] At some point, the binary

[82]Frank George, *Machine Takeover, The Growing Threat to Human Freedom in a Computer Controlled Society* (Pergamon Press, 1977) 6.

[83]'Krenn (n 29); John Zerzan, *Why hope?: the stand against civilization* (Feral House, 2015); John M Henshaw, *Does Measurement Measure Up? How Numbers Reveal and Conceal the Truth* (The Johns Hopkins University Press, 2006).

[84]Feenberg (n 39). Also, see Max Horkheimer, *Eclipse of Reason* (Oxford University Press, 1947, Continuum Publishing 1974, 2004).

[85]David M Berry, *The Philosophy of Software Code and Mediation in the Digital Age* (Palgrave Macmillan, 2011) 15.

[86]D Chandler, 'A World without Causation: Big Data and the Coming of Age of Posthumanism' (2015) 3 *Millennium: Journal of International Studies* 1.

[87]Mireille Hildebrandt and others, 'Introduction' *Digital Enlightenment Yearbook 2013* (IOS Press, 2013).

[88]TJ McIntyre and Colin Scott, 'Internet Filtering: Rhetoric, Legitimacy, Accountability and Responsibility', in Brownsword and Yeung (n 50) 109.

nature of Turing computation and its logical consistency eliminates any discretionary power as a capacity of the legal system to import extraneous knowledge to produce answers to the 'hard cases'.

As the consequences of such formalisation of reason, our aims and values like justice, equality, happiness, solidarity and tolerance, which have been inherent in or sanctioned by reason since the Enlightenment, lose their intellectual ground. Although such values exist in the constitutions of the sovereign states, they lack any confirmation by reason or agency to link them to an objective reality.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Emre Bayamlıoğlu is a doctoral researcher at Tilburg Institute for Law, Technology, and Society (TILT), Tilburg Law School, Tilburg University, Netherlands.

Professor Ronald Leenes is Director of Tilburg Institute for Law, Technology, and Society (TILT), Tilburg Law School, Tilburg University, Netherlands.

Chapter 3 (*Second paper*)

Contesting automated decisions: A view of transparency implications

• Emre Bayamlıoğlu, "Contesting automated decisions: A view of transparency implications" (2018), *European Data Protection Law Review*, 4(4), 433-446. https://doi.org/10.21552/edpl/2018/4/6.

In coming to terms with a difficult mathematical treatise or a poem written in a concentrated style, a magnifying glass will not help.[*]

[*] Valentino Braitenberg, *On the Texture of Brains: An Introduction to Neuroanatomy for the Cybernetically Minded,* Berlin Heidelberg: Springer-Verlag, 1977, 5.

Contesting Automated Decisions:

A View of Transparency Implications

*Emre Bayamlioglu**

This paper identifies the essentials of a 'transparency model' which aims to scrutinise automated data-driven decision-making systems not by the mechanisms of their operation but rather by the normativity embedded in their behaviour/action. First, transparency-related concerns and challenges inherent in machine learning are conceptualised as 'informational asymmetries', concluding that the transparency requirements for the effective contestation of automated decisions go far beyond the mere disclosure of algorithms. Next, essential components of a rule-based 'transparency model' are described as: i) the data as 'decisional input', ii) the 'normativities' contained by the system both at the inference and decision (rule-making) level, iii) the context and further implications of the decision, and iv) the accountable actors.

Keywords: Algorithmic Transparency, Automated Decisions, GDPR Article 22

The true nature of things may be said to lie not in things themselves, but in the relationships that we construct, and then perceive, between them.[1]

I. Introduction

Al-Jazari, a medieval engineer of mechanics and automata who was born in the town of modern day *Cizre (Turkey)*, Upper Mesopotamia, begins *The Book of Knowledge of Ingenious Mechanical Devices* by informing readers how the book had been ordered by his master, the King of *Diyar Bekir*, in sometime between 1204-1206 AD.[2]

> I was in his [the King] presence one day and had brought him something which he had ordered me to make. ... He said 'you have made peerless devices, and through strength have brought them forth as works; so do not lose what you have wearied yourself with and have plainly constructed. I wish you to compose for me a book which assembles what you have created separately, and brings together a selection of individual items and pictures'.

As Al-Jazari was valued not for his writings but rather his technical and mechanical talents, the King's order for a 'do-it-yourself' book may be seen as a straightforward expression of our everlasting desire to discover insights into how things work. But more importantly, the King also seems to have been aware that i) the mechanical knowledge of Al-Jazari and the relevant know-how were just as valuable as the actu-

DOI: 10.21552/edpl/2018/4/6

* Emre Bayamlioglu, is a researcher at the Tilburg Institute for Law, Technology, and Society (TILT) and also an external fellow of the Research Group on Law Science Technology & Society (LSTS) at Vrije Universiteit Brussels. For correspondence: <Emre .Bayamlioglu@uvt.nl>.

1 Terence Hawkes, *Structuralism and Semiotics* (2nd edn, Routledge 2003) 7.

2 Al-Jazari, *The Book of Knowledge of Ingenious Mechanical Devices:Kitáb fí ma'rifat al-hiyal al-handasiyya* (Translated and annotated Donald R Hill, Reidel Publishing Company 1974). The book contains detailed technical drawings and descriptions for the design, manufacture and assembly of Al-Jazari's allegedly constructed machines such as musical robot band, hand-washing automaton with flush mechanism, drink-serving waitress—which are regarded to be much beyond their time.

al machines he had constructed; and ii) insights into machines could not be obtained by cutting them open or dissecting into pieces, but rather required an 'explanation'. Even this almost arbitrarily chosen historical anecdote shows that, speaking of machines, transparency is not a degree of visibility, but rather, a capacity to see the relevant purpose(s)—not simply clearing of sight, disclosing, or making it open.

Put this way, the conundrum we face—namely, the transparency of 'data-driven automated decisions'[3]— is guided by the understanding that transparency cannot be seen as mere openness but presumes a communicative act, rather some form of information flow between the object and the observing subject. Following from this communicative dimension, transparency may also be regarded as a form of governance which may entrust subjects with more control over information. If implemented in the right context with ample instruments, transparency reduces the uncertainty and mitigates the effects of centralization, bias and information exclusivity. Yet, as a theoretical construct, 'transparency' cannot accurately conceptualise the information whose disclosure it hopes to prescribe.[4]

In the context of automated decision-making, approaches to transparency are rooted in the concern that as automated systems amass more data from an expanding array of sources, we end up delegating more power to machines to decide where and how we live, what we consume, how we communicate, how we are entertained, healed, and so on. This intensely algorithmic data-driven future transforms the marketplace, social relations and the relations with the sovereign as well as the very nature of the state. The capability of these systems to shape and influence the behaviour and the choices of individuals in pre-emptive and subtle ways further reinforce the view that those who control the algorithms 'permeate and exert power on all manner of forms of life.'[5] In this new narrative which has taken hold of the contemporary imagination, there is growing consensus that the lack of transparency cannot be resolved by standard disclosure practices, merely by 'throwing more information' at data subjects. Since data-driven modalities offer a multitude of opacities and informational asymmetries, the problem is not anymore what we can see but rather what we want know, and how much.

This article intends to identify the requirements of a transparency model which aims to explain automated decision-making systems not by the mechanisms of their operation but rather by the normativity embedded in their behaviour/action. The overall aim is to contribute to the legal scrutiny of automated decision-making by providing a synthesis of the implications of transparency as may be mandated by law (eg, the GDPR). Accordingly, Part II of the article systemises and provides a taxonomy of the transparency-related concerns and challenges inherent in machine learning (ML) under the banner of 'informational asymmetries.' This part explains how i) *(in)transparencies*, ii) *epistemological flaws* (spurious or weak causation), and iii) *biased processes* create cognitive obstacles on the side of the data subject in terms of contesting automated decisions. This macro-view reveals that the transparency implications of automated decision-making systems are too complex and dynamic to be addressed by merely remedying or sanctioning opacities. Part II concludes that automated decisions cannot be rendered reviewable, interpretable and thus contestable by opening up the 'black-box' but rather through a rule-based reconstruction of the decision-making process.

In Part III, essential elements of a transparency model for the effective contestation of automated decisions are formulated as: i) the data as 'decisional input/cues'; ii) the 'normativities' contained both at the inference and decisional (rule-making) level; iii) the

3 A*utomated* and *data-driven* are two different concepts. In the literal sense, alarm clock set to ring at 07:00 AM every day is perfectly automated but not data-driven. On the other hand, the refrigerator with a thermostat is both data-driven and automated. The question arises whether there could be systems that are data-driven but not automated. Even if this is rarely the case, the answer is in principle affirmative. Early judicial aids for sentencing could be regarded as data-driven or statistics-based but still not automated in that the human judge made the final decision. Therefore, throughout this article, we prefer to use the variations of the generic term 'automated data–driven decision-making', and avoid the term 'algorithmic' except for the ease of reference, or where the usage or context necessitates (eg, *algorithmic transparency*, *algorithmic scrutiny*).

4 Mark Fenster, 'Transparency in Search of a Theory' (2015) 18 European Journal of Social Theory; Christopher Hood, 'Transparency in Historical Perspective' in Christopher Hood and David Heald (eds), *Transparency: The Key to Better Government?* (OUP 2006); John Roberts, 'No One is Perfect: The Limits of Transparency and an Ethic for 'Intelligent' Accountability' (2009) 34 Accounting, Organization and Society 957; Andrea Brighenti, 'Visibility: A Category for the Social Sciences' (2007) 55 Current Sociology 323; Ida Koivisto, 'The Anatomy of Transparency: The Concept and its Multifarious Implications' (EUI Working Papers, 2016); Hans Krause Hansen and Mikkel Flyverbom, 'The Politics Of Transparency And The Calibration Of Knowledge In The Digital Age' (2014) 22 Organization; David Heald, 'Varieties of Transparency' in Hood and Heald (n 4).

5 Andrew Iliadis and Federica Russo, 'Critical Data Studies: An Introduction' [2016] 3 Big Data & Society.

context and further implications of the decision; iv) the accountable actors. The model aims to construct a link between the data as input and the ensuing effects within a normative framework. It is guided by the premise that '[...] when demanding explanation from humans, what we typically want to know is whether and how certain input factors affected the final decision or outcome.'[6]

Part IV points to the possible implementation problems at the technical, economic and legal level, revealing that the transparency needs of an effective contestation scheme go much beyond the disclosure of algorithms or other computational elements. That is, a viable option would need to strike a balance between the social and political concerns, the legal limits and the technological affordances, while taking into consideration the actual functioning of specific devices and data-driven business models.[7] The final part synthesizes the findings and concludes with the formulation of the preliminaries of what could be called a 'contestability scheme'. The scheme thus provided may be seen as a theoretical guide for compliance with transparency obligations such as those provided under the EU data protection regime (the GDPR).[8]

II. Informational Asymmetries in ML-Based Decisions

> *[...] data-driven analytics go beyond the limits of the known and seek to unveil and rationalize the unknown. Not only do they seek to render the future actionable, they also promise to provide a glimpse into the future by creating a new and distinct form of knowledge about it.*[9]

Automated decision-making processes are characterised by different types of 'informational asymmetries'[10] between the system and the affected individual. Increasingly complex and adaptive properties of these systems render their technical dimension and inner workings opaque to human cognition in that a *prima facie* analysis of the data and the algorithms may not have any plausible link to the result.[11]

For the purposes of contestation, being *inscrutable* (opaque) and being *unpredictable* are different types of informational asymmetries giving rise to distinct cognitive deficits. First, the complexity of the operations might make a mental follow up of the decision very burdensome, if not impossible. Even in case of allegedly less complex *decision tree* algorithms (which may be seen as a well-specified mechanised process), the size of the model (total number of nodes or branches) may grow much faster than the time needed to perform inference.[12] Second, the adaptive and dynamic nature of the data-driven systems create obstacles regarding the foreseeability of the result. The unpredictability of the outcome could be attributed to the systems' capability to modify their responses according to the changes in the environment. An adaptive algorithm (also known as 'non-deterministic') may produce different results for each instant of its execution. Hence, while complexity can only be a barrier to big-picture understanding—not to understanding which factors might have changed a particular outcome— adaptive algorithms seriously diminish the chance of predicting the results for a particular set of input.[13]

Commonly used pre-emptive strategies and invasive models, where future is anticipated and acted upon, add a separate layer of opacity resulting in a 'mental invisibility' on the side of the individuals subject to automated decision. Pre-emptive strategies aim to reduce risks by formulating a default 'reason-

6 Finale Doshi-Velez et al, 'Accountability of AI under the Law: The Role Of Explanation' (Berkman Klein Center for Internet & Society working paper, 2017) <http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372584> accessed 29 November 2018.

7 Giovanni Comandé, 'Regulating Algorithms' Regulation? First Ethico-Legal Principles, Problems and Opportunities of Algorithms' in Tania Cerquitelli, Daniele Quercia and Frank Pasquale (eds), *Transparent Data Mining for Small and Big Data* (Springer 2017) 169-206, 190.

8 This article is a part of a PhD project which aims to clarify whether the specific provisions of the GDPR accommodate the below defined transparency requirements. Taking the model in this paper as the benchmark of transparency, the thesis inquires whether the GDPR provides the necessary normative arsenal for the effective contestation of automated decisions. The research further extends to the question whether (or to what extent) Intellectual Property rights as referred in the GDPR stand as a legal impediment for the implementation of the model.

9 Matthias Leese, 'The New Profiling: Algorithms, Black Boxes, and The Failure of Anti-Discriminatory Safeguards in The European Union' (2014) 45 Security Dialogue 494, 501.

10 Bruno Lepri et al, 'The tyranny of data? The bright and dark sides of data-driven decision-making for social good' (2016) arXiv:1612.00323.

11 Jenna Burrell, 'How the machine 'thinks': Understanding opacity in machine learning algorithms' (2016) Big Data & Society 3(1) 1–12.

12 Zachary C Lipton, 'The Mythos of Model Interpretability'(2016) arXiv:1606.03490v3 accessed 29 November 2018.

13 Edward Felten, 'What does it mean to ask for an 'explainable' algorithm?' (*Freedom to Tinker*, 31 May 2017) <https://freedom-to-tinker.com/2017/05/31/what-does-it-mean-to-ask-for-an-explainable-algorithm/> accessed 29 November 2018.

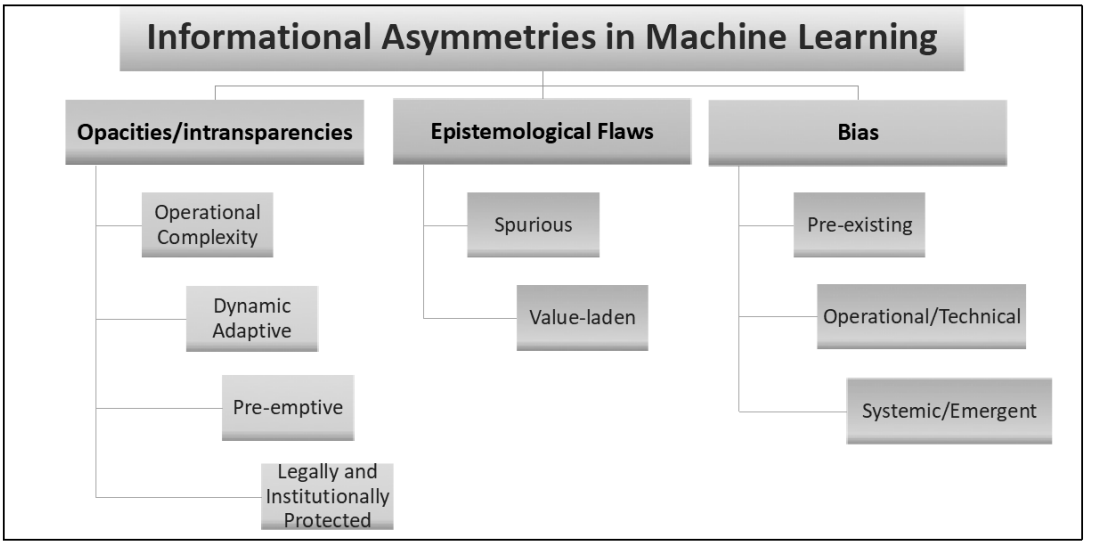

Figure 1. *Informational asymmetries in machine learning*

able' conduct—striving to act before the identification or even the formation of a determinate threat.[14] The effect is so subtle because the desired behaviour is induced through the elimination of choices that are therefore not revealed to the data subject. Preemptive systems do not merely maintain an 'awareness', they are also reflexive of the feedback they receive, entrapping individuals in a cognitive environment where they are constantly implicated.[15] Especially in ambient intelligence environments, the system senses, analyses and models individuals by anticipating their possible behaviour in order to preempt what is deemed inappropriate without conscious mediation.[16] Hence, comprehensible information about such strategies is an essential marking point in determining the boundaries and the context of any decisional framework.

Legal and institutional impediments as a source of lack of transparency in algorithmic processes manifest themselves either as a culture of confidentiality and secrecy promulgated by the businesses, governments or other organizations of interest, or in the form of legal claims primarily based on intellectual property rights and in particular trade secrets. The secrecy serves a dual purpose, both to prevent gaming behaviour by the data subjects and to exclude rivals. This embedded reflex of corporations and institutions obscure not only the inner workings of the systems but also other relevant and probably critical legal, economic and political contingencies.

Above intransparencies conceal and amplify further informational asymmetries in the form of *epistemological flaws*, and *biased processes* inherent to ML. The basic epistemological impediment is the difficulty of telling the difference between events that are causally related and events that are merely associated (correlative) with each other in time or place. Although it is paramount to humans to understand the reason (mechanism) behind the associations one encounters in the real world, most of the big data practices focus on the potential exploitative and invasive uses of data, rather than the nature and the

14 Leese (n 9) 498, citing Ben Anderson, 'Preemption, precaution, preparedness: Anticipatory action and future geographies'(2010) Progress in Human Geography 34(6) 777, 792.

15 Felix Stalder, 'From inter-subjectivity to multi-subjectivity. Knowledge claims and the digital condition' in Emre Bayamlıoğlu et al (eds),''BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen' (AUP 2018).

16 Simon Elias Bibri, *The Human Face of Ambient Intelligence: Cognitive, Emotional, Affective, Behavioral And Conversational Aspects* (Atlantis Press 2015).

quality of the inferred knowledge.[17] Neglecting experience and intuition, decision-making becomes increasingly based on finding correlative patterns. These possibly 'blind correlations' do not stem from predefined hypotheses and therefore do not conform to the principle of cause and effect. Moreover, 'causation' has no explanatory link unless narrated through a theory and implemented as a model based on that theory. Causality, especially in data-driven practices, even where it is 'properly' established, is a question of model-building. In any knowledge query, the model—together with the heuristics it embodies— is important to make sense of the world and thus, to extrapolate beyond the inherent constraints of the observed domain.[18] Put in other words, there is no information in the data about a possible causal relation between the weather and the ice cream sales. Apparently, we take for granted the underlying heuristics, domain knowledge and the common sense that we employ to fill the gap between the evidence we have and the inference we make.[19] That is, what people ate has no bearing on the weather, and the sun will rise tomorrow even if the rooster remains silent (or, we just don't know).[20] While correlations appearing in large datasets may possess insightful value from a certain perspective, they may simply turn out be spurious as mere coincidences when the causal model is changed.

Lastly, bias in ML processes is another dimension of automated data-driven decisions which stands out as a source of informational asymmetry. Independent of discrimination, bias in data-driven decision-making systems refers to an inclination or an outlook to present or hold a partial perspective, including the refusal or the ignorance to consider other possible aspects. Bias explains any tendency and interest of a system to act in a certain way or to yield certain results. An algorithm is not merely a neutral transformer of data, every algorithm which somehow aims for sorting or prediction will eventually prioritise certain criteria and establish some kind of ranking.[21] As primarily a computational and data-originated problem—albeit with strong intertwined economic, political and social roots and underpinnings—bias in automted decisions may be studied under a tripartite categorisation: *pre-existing bias, technical/operational bias* and *systemic/emergent bias*.[22] However, such categorisation may not be taken as establishing distinct compartments of analysis. Almost in every case, bias is a fusion of various dynamics conflated in a complex way, based on the approach chosen to model the problem at hand. Every stage of ML has a direct or indirect bearing on the final interpretation. Different stages or components of big data analysis cannot be analysed in isolation but rather require a systemic review. As bias is determined by the entire process of decision making, it cannot be detected simply by analysing the end result. Bias emerging throughout the data collection and analysis stages may or may not translate into undesirable discriminatory results at the final interpretation/decision stage. It may also be the case that different types of biases may offset each other, and at times, designers may use this as a strategy for the elimination of bias originating from the training data—eg by employing an algorithm which ignores the biased part of the data. Bias is a polymorphic and contextual concept with many facets and dimensions, and we cannot *per se* conclude that all bias is harmful and must be outlawed categorically. There are many different types of bias depending on the standard being used.[23] Irrespective of discriminatory effects, not being able to understand whether and why a system innately leans towards certain outcome severely impairs the objective interpretability of the decision. This deprives the data subjects of some important information which is essential to argue against a specific decision.

Independent of harms that may be addressed by the law (eg, unfair results and procedures, invasive practices, exclusionary market control, and further

17 David Chandler, 'A World without Causation: Big Data and the Coming of Age of Posthumanism' (2015) Millennium: Journal of International Studies 1–19, 2; Thomas W Simpson, 'Evaluating Google As An Epistemic Tool' in Harry Halpin and Alexandre Monnin (eds) *Philosophical Engineering Toward a Philosophy of the Web* (Wiley Blackwell 2014) 97-116; Rob Kitchin, 'Big Data, new epistemologies and paradigm shifts' (2014) Big Data & Society 1–12, 4.

18 Mireille Hildebrandt and Katja de Vries (eds), *Privacy, Due Process and the Computational Turn* (Routledge 2013); John H. Holland, *Hidden Order: How Adaption Builds Complexity* (Basic Books 1995) 5.

19 Peter Lipton, *Inference to the Best Explanation*, (1991 1st edn, Routledge 2004) 7.

20 Judea Pearl and Dana Mackenzie, *The Book of Why, The New Science of Cause and Effect* (Basic Books 2018) 3.

21 Scott J Muller, *Asymmetry: The Foundation Of Information* (Springer-Verlag 2007).

22 Batya Friedman and Helen Nissenbaum, 'Bias in Computer Systems' (1996) 14 ACM Transactions on Info Sys 330.

23 David Danks and Alex John London, 'Algorithmic bias in autonomous systems' International Joint Conference on Artificial Intelligence (IJCAI) 2017.

threats to self-autonomy), the above taxonomy (summarised in Figure 1) is an attempt to systemise and theorise potentially problematic dynamics inherent to ML. It cannot be taken as a blueprint or some sort of 'one fits for all' template for investigation, but rather offers a conceptual mapping with legally meaningful pointers.[24] Notwithstanding the legal, technical or epistemological dimension, each informational asymmetry renders individuals prone to manipulation and exploitation in a specific manner, while also potentially incapacitating them from appealing against unlawful or somehow undesirable results. In light of these intertwined, elusive, constantly co-opting and overlapping dynamics, one should not understand contestation as challenging of a certain result but rather as making the entire decision-making process reviewable and interpretable through concrete transparency requirements.

III. Transparency Requirements to Contest Automated Decisions

1. Normativity: The Key to Theorising Transparency

> *[...] algorithmic decision-making necessarily embodies contestable epistemic and normative assumptions.*[25]

Following from the above, the transparency implications of automated decision-making systems are too complex and dynamic to be addressed by merely remedying or sanctioning the informational asymmetries. In an environment of constant data flux, the person who is target of the decision may not be capable of mapping a particular outcome (eg, a classification) against a given input. Speaking of scrutiny of automated decisions, theorising transparency with a view to see the entire system 'at work' is a territory ever expanding as we attempt to map it. The opacities and informational asymmetries inherent in ML results in a 'mental invisibility' that may only be counteracted through a visibility of a different type. For the purposes of contestation, such as the one provided under Article 22 of the GDPR, this entails an 'actionable transparency', an instrument to an effective and practical enforcement of rights.[26]

As decision-making systems are goal-oriented, their behaviour may eventually be attributed to the inherent values and assumptions which guide their response to a given input—allowing us to expect a related 'normativity' in the system's output.[27] Since, by themselves, facts (data) cannot provide reasons for action[28], looking through the lens of normativity may inform us about the motives, assumptions and the further decisional criteria underlying the system.

Accordingly, challenging the truth claim or the accuracy of a decision, thus contesting 'what ought to be' in a given situation, will initially require a conceptualisation of the outcome as the result of a 'rule-based' process where certain input is rightfully matched with certain results— akin to a legal system where rules (norms) are applied to facts (input data) to make decisions (output data). Hence, a 'rule-based explanation' means that given certain input data, the decision (output data) should be interpretable, verifiable, and thereby contestable with reference to the rules (normative framework) that are operational in the system. In the context of automated decisions based on profiling, this would refer to how and why a person is classified in a certain way, and what consequences follow from that.[29]

---

24 For similar conceptualisations see, Brent Mittelstadt et al, 'The ethics of algorithms: Mapping the debate' (2016) Big Data & Society 10.1177/2053951716679679; Martijn van Otterlo, 'Gatekeeping Algorithms with Human Ethical Bias' (2018) arXiv:1801.01705v1.

25 Reuben Binns, 'Algorithmic Accountability and Public Reason' (2017) Philosophy & Technology. Also, see Mireille Hildebrandt, 'The New Imbroglio. Living with Machine Algorithms' in Liisa Janssens (ed), *The Art of Ethics in the Information Society. Mind you* (AUP 2016).

26 Mireille Hildebrandt, 'Privacy as Protection of the Incomputable Self: Agonistic Machine Learning' (2019 forthcoming) 19(1) Theoretical Inquiries of Law <http://dx.doi.org/10.2139/ssrn.3081776>.

27 Stefano Bertea, *The Normative Claim of Law* (Hart Publishing 2009) 11-12; Joseph Rouse, 'Social practices and normativity' (2007) 37(1) Philosophy of the Social Sciences 46; M Franssen, 'Artefacts and normativity' in A Meijers (ed), *Handbook of the philosophy of science: Vol. 9: Philosophy of technology and engineering sciences* (Elsevier 2009); George Pavlakos and Veronica Rodriguez-Blanco, *Reasons and Intentions in Law and Practical Agency* (CUP 2015); MJ Vries, SO Hansson and A Meijers (eds), *Norms in Technology* (Springer 2013); Michael Giudice, *Understanding the Nature of Law. A Case for Constructive Conceptual Explanation* (Edward Elgar 2015).

28 Joseph Raz, *The Authority of Law* (Clarendon Press 1979); Maarten Franssen, 'The Good, the Bad, the Ugly... and the Poor: Instrumental and Non-instrumental Value of Artefacts' in P Kroes and P Verbeek (eds), *The Moral Status of Technical Artefacts*, (Springer Science+Business Media 2014). Also see, J Dancy, 'Non-naturalism' in D Copp (ed), *The Oxford handbook of ethical theory* (OUP 2006) 122.

29 Ronald Leenes, 'Reply: Addressing the Obscurity of Data Clouds' in Mireille Hildebrandt and Serge Gutwirth (eds), *Profiling the European Citizen: Cross-disciplinary Perspectives* (Springer 2008).

Such modelling which maps input with the effects/consequences within a normative framework helps contextualise the decision at the appropriate level of generality for the purpose of constructing a domain-agnostic contestation scheme. Understanding the potential normative impact of the data-driven decisions requires investigating what types of behaviours or constraints are enforced or inhibited by a particular set of input (data).[30] The concrete transparency requirements of such a rule-based model (RbM)— as an operable scheme— entail explanations, disclosures, verifications and justifications with regard to the below aspects and components of the system.

2. Data Features as Decisional Cues: Data as the Representation of the World

> *[...]decision-making processes that infer new insights from data, use these insights to decide on the most beneficial action, and refer to data and an inference process to justify the chosen course of action.*[31]

The rendering of automated decisions transparent cannot mean the disclosure of data as a mere list but in a structured and functional way, as part of the process where given inputs produce a specific outcome. Thus, rather than access to every type and instance of data, RbM seeks for an understanding of how the data is being translated into 'input' for the system.

To solve a problem, it is necessary that a computation manipulates a representation of something meaningful in the real world. 'Meaning of a computation depends on the meaning of the representation it transforms.'[32] Hence, any normative contestation will initially need the knowledge of what the system relies upon in order to reach results. This requires a perspective which treats the concept of 'data' not as a tool of insight, but simply as 'factual' or representational input for the purpose of inference.

In a ML process, data instances exist as values of feature variables where each feature such as age, height and weight is a dimension of the problem to be modelled. Depending on the nature of the analysis and the type of data available, features may also contain more constructed and computed representations such as one's habit of eating deep-fried food, educational level, or speaking a dialect. Features as decisional cues refer to the totality of the relevant data representations extracted from a larger set of feature variables.

The objective of a ML process is the identification of statistically reliable relationships between the feature variables and some target variable (eg healthy or not, or at least 70% healthy). In case of personal data processing, a feature space maps how people will be represented as inputs to the algorithm. The features that a system regards to be significant and their relevant weightings help us understand which inputs (inferences) factored into a decision to get to the final result. In broader terms, we need to know how the system relates to the world. What it really 'learns' with regard to persons, places, things and their actual existence. Although, with sufficient data, it is possible to construct a predictive model of, for example, eating or driving habits, the fact remains that data features and the ensuing inferences are only part of an array of possible ways of defining what a careful driver or healthy eater means.

Selecting a subset of relevant features that best correlate to the target variable of interest is not a purely empirical process but one constrained by the available data, and guided by design and implementation choices of the system. 'Determining which features should be considered is part of the determination of how the decision should be made; representing those constructs in measurable form is a separate and important step in the process.'[33]

3. The Normative Grounds

a. Two Tales of Normativity

A norm can be understood as an expected pattern of behaviour which imposes constraints on human or

30 Mireille Hildebrandt, 'A multifocal view of human agency in the era of autonomic computing' in Mireille Hildebrandt and Antoinette Rouvroy (eds), *Law, Human Agency, and Autonomic Computing: The Philosophy of Law Meets the Philosophy of Technology* (Routledge 2011) 2.

31 Patrick Allo, 'Mathematical values and the epistemology of data practices' in Bayamlıoğlu et al (n 15).

32 Martin Erwig, *Once upon an Algorithm: What Stories Can Teach Us about Computation* (MIT Press 2017) 50.

33 Sorelle A Friedler, Carlos Scheiddegger and Suresh Venkatasubramanian, 'On the (im)possibility of fairness' (2016) arXiv:1609.07236v1.

non-human subjects.[34] Statements, processes or mechanisms are regarded to be normative when, rather than articulating *what it is*, they claim to influence or dictate how certain things or certain matters should be.[35] In this general perspective, contestation of automated decisions can be based on two grounds, scrutinising two different types of 'normativity'.

First, decisions may be contested on the basis of the selection and construction of the relevant data features that the decision relies upon (rules for constructing decisional cues—ie inference rules). What is questioned here is whether inferences made by way of selected features are sufficiently informative and causally reliable for the given purpose, eg whether one's search for deep-fryers over the internet suffices for the inference of her/his eating deep fried food, and consequently being classified as risky. The normativity of decisional cues (selected features) lies in their being formal constructions by way of if-then rules. Both the accuracy and suitability of the features together with the methodology used for their selection and construction could be subject to normative scrutiny.

Second, normativity operates as a set of rules which describe how a certain ML outcome (target value) is translated into concrete results in a wider decision-making framework, eg a certain health risk resulting in an increased insurance premium. Based on complex relational conditions, there may be numerous if-then rules embedded in the functioning of the system.[36] What is important to note: not all decisions are necessarily yes/no type. The decision may be of non-binary nature which ends up with several categories or rankings each varying in consequences and scope, eg *recruit, ignore,* or *save as reserve* in CV filtering. The question is: what is the meaning of the target variable(s) obtained? What score (in numeric or other quantified form) would suffice, for instance, for a successful loan and most importantly why? Decisional norms are shaped by the hypotheses and assumptions about the root cause of the targeted problem—eg, avoiding customer churn, better distributing insurance risks. This type of scrutiny eventually reaches back to the goals and values encoded in the system, together with the underlying justifications and ratiocinations.[37]

Take the example of a political micro-targeting campaign where speech analysis can detect one's accent or dialect to predict her/his political opinion. Irrespective of legal or ethical admissibility of such inquiry, dialect may be regarded as a factual input the accuracy or the validity of which may be challenged on empirical basis. On the other hand, the selection of the 'suitable' political content based on this 'factual' finding is the result of the decisional criteria contestation of which would require a different reasoning and argumentation.

b. Epistemological Gaps

> *Data-driven architectures operate at another level that sublimates rather than externalizes the normativity that directs and coordinates our interactions.*[38]

Both the determination of the decisional cues and the ensuing results are normative undertakings which may be reconstructed in the if-then form (if condition 1 ∧ condition 2 ∧ condition 3, then outcome). Thus, theoretically every decision can be decomposed to find out which rules have been followed in what order. However, in case of automated decisions, neither the inferences nor the decisional norms that produce the outcome reveal themselves easily. Problems are not always as straightforward or simply verifiable as is the relation between eating habits and increased health risk —a plausible assumption based on common sense or past data.

In most of the cases, decisional cues (input) do not exist as readily available features but they need to be constructed from a multi-dimensional data set. As feature space becomes high-dimensional (meaning that a great many variables are repeatedly correlated), this entails that features are further selected and extracted to reduce the dimensionality of the data and consequently, the complexity of the model. Feature selection (FS) means choosing of the best possi-

34 Mireille Hildebrandt, 'Technology and the end of law' in E Claes, W Devroe and B Keirsbilck (eds), *Facing The Limits of The Law* (Springer 2009).

35 Sylvie Delacroix, *Legal Norms and Normativity: An Essay in Genealogy* (Hart Publishing 2006) xi.

36 Lucas Introna, 'Algorithms, Performativity and Governability' (2013) Governing Algorithms: A Conference on Computation, Automation, and Control, New York University.

37 John Zerilli et al, 'Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?' (2018) Philosophy & Technology <https://doi.org/10.1007/s13347-018-0330-6>.

38 Mireille Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Edward Elgar Publishing 2015) xiii.

ble subset of features from the feature space where feature extraction (FE) converts multi-dimensional data to a lower-dimensional space in a reduced form. While FS results in a more interpretable set of relevant features, in case of FE, the link with the reality may get lost in the mathematics and physical meanings of features may not be retained—rendering it impossible to clarify how the final output relates to any specific feature.[39] The result is a set of overly constructed and computed features where correlations between feature variables and the target variable do not depend on the conventional understanding of 'cause and effect'— introducing seemingly irrelevant input. Think of, for example, using spelling mistakes for predicting overweight persons in health insurance, or the length of the screen name of one's social media account for credit scoring. This implies that the assumed link between the input and the actual behaviour may not only turn out to be intrusive, incorrect, or invisible, but may even be non-existent due to spurious correlations. Especially in case of deep learning models using neural network type of algorithms, normative scrutiny of these overly constructed features may not be possible primarily because these systems have not been designed with such an assessment in mind. This is best explained by Dormehl in the own words of a system designer:[40]

> Berk [the system designer] makes no apology for the opacity of his system. 'It frees me up,' he explains. 'I get to try different black boxes and pick the one that forecasts best.' What he doesn't care about is causal models. 'I make no claims whatsoever that what I'm doing explains why it is that individuals fail,' ... 'I'm not trying to develop a cause-and-effect rendering of whatever it is that's going on. I just want to forecast accurately.'

As automated decision-making systems rely upon patterns found in data even when such inferences seem unwarranted, this brings the possibility that individuals might be judged for activities that are associated with particular racial, ethnic, or socioeconomic groups — exacerbating existing forms of bias and inequalities. What is more, ML systems classify individuals at an aggregate level based on unknown or unintended commonalities. The groups thus created, might not be easily definable, or even recognisable due to their seemingly random nature.[41] For example sorting one's Facebook friends through a clustering algorithm may discover a group such as 'church friends' though the user has never contemplated such a grouping of her/his social media contacts.[42] Rather than being based on factual representations, categorising through data could be seen as procedures that initially create the groups they aim to define.[43] The seeming neutrality of data somehow naturalises this segmentation, and falsely renders its own construction—or say, normativity— invisible as a regulatory process. Different modes of segmenting populations may be seen as a means to create new criteria both for the identification of the predictive targets and for the application of the rules.

4. The 'Context' of the Decision

There is nothing either good or bad, but thinking makes it so.[44]

Seen as techno-regulatory processes, automated decision-making systems have different modes of interaction with the physical world, and thus can influence the real-life situations in an array of ways. To fully evaluate the automated decisions for the purposes of contestation, the context of the decision—as the particular situation, environment or domain in which the decision is to be made—is a key piece of information. The knowledge of the context enables both the evaluation of the decisional framework and the decision alternatives with a projection of further effects. For the purposes of contestation, context is an additional complexity in that a decision may be 'good' in a particular context but less 'good' in other contexts.[45]

39 Li Jundong et al, 'Feature Selection: A Data Perspective' (2017) ACM Computer Surveys 50(6) Article 94 <https://doi.org/10.1145/3136625>.

40 Luke Dormehl, *The Formula: How Algorithms Solve All Our Problems-And Create More* (Perigee Books 2014) 128.

41 Anton Vedder, 'Why data protection and transparency are not enough when facing social problems of machine learning in a big data context' in Bayamlıoğlu et al (n 15).

42 Motahhare Eslami et al, 'Friend Grouping Algorithms for Online Social Networks: Preference, Bias, and Implications' in LM Aiello and D McFarland (eds), *Social Informatics* (Springer 2014).

43 Karoline Krenn, 'Markets and Classifications - Constructing Market Orders in the Digital Age. An Introduction' (2017) 42(1) Historical Social Research.

44 William Shakespeare, *Hamlet, Prince of Denmark*, Philip Edwards (ed) (CUP 2003) 141.

45 Zhiwei Zeng et al, 'Context- based and explainable decision making with argumentation.' (IFAAMAS, 2018) in AAMAS-18 (submitted).

For a normative assessment, the outcome in ML would need to be nested within a larger decision-making model. Thus, any reason-giving (explicatory) transparency approach would naturally entail comprehensible and verifiable information about the context wherein the decision is made and implemented. Understanding the context of the decision is not confined to an evaluation of the end result but also extends to identifying the reliability and validity issues as well as the operational limits regarding the collection and transformation of the data. As necessary transparency information, the context provides an assessment of the model together with the further impact of the decision in light of the declared and undeclared purposes of the system.

The knowledge of the context primarily involves informing of the data subject about where the decision starts and ends. Whether the system interoperates with other data processing operations, which other entities and authorities are informed of the decision, and for what other purposes and in which other contexts the results could be used are also crucial information. For example, credit scoring could be used as a proxy for other types of risks such as insurance claims, workplace trustworthiness, rent payment, telecommunications, or even utilities pricing.[46] This type of reutilisation of the data across different economic segments is a source of additional concern since sharing between industries may lead to exclusion also from nonfinancial services and thus to wider economic discrimination.

Lastly, understanding the context of the decision, with contestation in mind, requires not only the knowledge of why a decision was made but also why a different decision was not made.[47] Asking the question, for instance, what is qualified as healthy eating could reveal the weakness of the causal model while also exposing the flaws in the normative set-up. The conditions under which the decision would not have been made is a type of information necessary to construct counterarguments. Considering the epistemological limitations (causal weakness) of correlations due to their possibly spurious nature, contrastive explanations (counterfactuals) as to the differences between two decisions enable a meaningful distinction between several decision alternatives:

> Counterfactuals are the building blocks of moral behavior as well as scientific thought. The ability to reflect on one's past actions and envision alternative scenarios is the basis of free will and social responsibility. The algorithmization of counterfactuals invites thinking machines to benefit from this ability and participate in this (until now) uniquely human way of thinking about the world.[48]

5. Agency (Responsible Actors) behind the Automated Decisions

> *[...]only those who will stand behind their actions should exercise authority.*[49]

Information about the responsible actors is an essential element of an actionable transparency model, meaning that the implications of automated decisions must be situated and analysed in an institutional framework. The impact of the automated decisions may not be properly contextualised without knowledge of the commercial or other institutional interests underlying the process. Lacking this particular dimension, the transparency model remains incomplete.

With a rising dynamism in the data industry and the pervasion of big data technologies in general, we see a proliferation of the actors interacting with one another.[50] The 'agency' behind automated decisions is not monolithic but often related to a plethora of conflicting, competing and partially overlapping interests and objectives which are linked to multifarious commercial frameworks and stately functions.

Contestation of automated decisions may be based on different grounds such as decisional norms, inferences, errors in data, or the accuracy of calculation. These different types of contestation may relate to several actors who may have conflicting concerns

---

46 Danielle Keats Citron, 'Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age' (2007) 80 S Cal L Rev 241, 295; Federico Ferretti, 'The Legal Framework of Consumer Credit Bureaus and Credit Scoring in the European Union: Pitfalls and Challenges— Overindebtedness, Responsible Lending, Market Integration, and Fundamental Rights' (2014) XLVI Suffolk University Law Review 791, 823 <http://suffolklawreview.org/wp-content/uploads/2014/01/Ferretti_Lead.pdf> accessed 29 November 2018.

47 'To explain why P rather than Q, we must cite a causal difference between P and not-Q, consisting of a cause of P and the absence of a corresponding event in the history of not-Q.' Lipton (n 19) 41. Also see Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences' (v2, 24 May 2018) arXiv:1706.07269v2

48 Pearl and Mackenzie (n 20) 10.

49 James Grimmelmann, 'Regulation by Software' (2005) 114 Yale LJ 1719, 1734.

50 Stalder (n 15).

with regard to transparency. Hence, identifying the transparency requirements of a specific type of scrutiny also depends on the composition of the actors behind the decision.[51] This requires a purposeful mapping of the institutional structures and the intricate web of relations among those who may be responsible for different parts or aspects of a decision (ie the data brokers, public and private clients, service providers, regulators, operators, code writers and system designers).

Along with this, in techno-regulatory settings, we increasingly observe a take-over of the regulatory functions—once carried out by public authorities—by private interests. These hybrid and fluid structures controlling digital platforms and data-driven systems are of special concern due to the difficulties in identifying the accountable parties as well as the essential processes with regard to a specific decision.[52] Leaving aside their monopolistic power, platforms such as Google act less like a private company when they cooperate with governments in the field of law enforcement, eg scanning e-mails for evidence of child pornography. Unlike state-backed modern legal systems, in the regulatory sphere of automated decisions, there is no sovereign that is responsible for rule-making, administration and adjudication through separate and clearly distinct state functions. Although new algorithmic decision-makers may seem sovereign over important aspects of individual lives, they mostly assume the role of unaccountable intermediaries.[53]

IV. Implementation of the Model: Current Impediments, Future Horizons

The above transparency model and the following informational requirements form an abstract template which aims to systemise certain core elements of automated decision-making systems. Each component or dimension is a content-agnostic formulation to facilitate both a normative review of the system and the contestation of a specific decision. They may not be seen as independent assessment criteria but rather need to be implemented and put to use in an interdependent way. What is intended here is a 'legal reading' of a technology with unprecedented regulatory capacities in direct, indirect and subtle ways.

The implementation of the model does not necessarily bring forward a full disclosure of the entire set of elements or structures comprising the system. This is neither legally possible nor technically or economically feasible since data-driven systems are too sophisticated to know everything about them. As such, the transparency information(requirements) entailed by the above model does not pose a direct threat to the 'secret sauce' of data 'wizards' which may be subject to trade secrecy claims.[54] As a matter of fact, the possible complications arising from the implementation of a rule-based model are much beyond the problem of disclosure of secret algorithms or the proprietary code embodying them. The knowledge of the inferences and the normative grounds underlying a specific decision or a decision-making system together with a mapping of the stakeholders involved, could uncover several critical specifications or the essential properties of the system. Therefore, rather than opening up the 'black-box', a more substantial resistance to such models may arise due to the exposure of data controllers' commercial logic, profit-maximising strategies, and other possibly deplorable conduct in a bird's-eye view.

As a theoretical construct, the RbM and the ensuing transparency requirements draw the horizon of the desirable (but not necessarily the possible or the optimal) for contestation purposes. Various concerns may be raised with regard to the economic, technical or legal permissibility of these requirements. A viable implementation of the RbM entails the consideration of certain impediments inherent in the process, namely i) the legal limits: security/integrity and secrecy; ii) the technical constraints, eg inscrutable algorithms; and iii) the economic feasibility. As such, the implementation of the model at various levels through different tools and modalities needs to strike a balance among the risks, computational difficulties and the economic restraints while taking into account the legal limits—eg, to prevent competitors from reverse-engineering the scoring model or customers from

51 Adrian Weller, 'Challenges for Transparency' (2017) International Conference on Machine Learning (ICML) Workshop on Human Interpretability <https://arxiv.org/abs/1708.01870> accessed 29 November 2018.

52 Kenneth A Bamberger, 'Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State' (2006) 56 Duke LJ 377, 386-387.

53 Danielle Keats Citron and Frank Pasquale, 'The scored society' (2014) 89(1) Washington Law Review 1–33, 19.

54 'Our secret sauce' (*Official Google Blog*, 25 February 2008) <https://googleblog.blogspot.nl/2008/02/our-secret-sauce.html> accessed 18 December 2017.

gaming the smart grid.[55] Depending on the object and focus of the analysis together with its scope, intensity and duration, the necessary forms and degree of transparency (eg notification/ disclosure, algorithmic audit and *legal protection by design* principles) cannot be detailed in the abstract, but require a further refinement in light of the specificities of the domain, and the context of the data operations in question.

So long as certain contestability standards are not *ex-ante* imposed by a 'legal protection by design' approach, there will always be instances where the normativity implicit in the system could not be articulated by a review of the system in the abstract.[56] This is primarily because adaptive systems operate on dynamic correlation patterns where the decisional rule may itself emerge autonomously from the streaming data. The 'norm' is no longer predetermined, but constantly adjusted. Such *fluid hypotheses*[57] make any challenge on normative credentials of the system hard to formulate and thus, the decisional criteria remain vague and cannot be pinned down in sufficient precision.[58]

Despite these impediments, there are various efforts to develop methodologies and software tools for explaining the so-called 'black-box' models. An example of such efforts which is compatible with the idea of normative scrutiny is the LIME project which aims to disclose the implicit rules behind ML-based predictions.[59] The project develops an interpretable model taking on the predictions of a supposedly uninterpretable (black box) model.[60] These technical solutions decompose predictions on the basis of the contribution of each attribute/feature in the result.[61] The tools for this purpose generally focus on importance-measuring methods that operate on the individual level, explaining what most important variables were for a specific result. Differently, approaches at the system level (algorithm-wide) explore how important were the variables to the algorithm's training.[62] Success of these technical solutions both for systemic review and for the scrutiny of specific decisions will require informed choices at the model selection and implementation stage of the automated decision-making systems.

None of the current transparency tools can take into account the context and environment in which the decision is made, lacking any guidance on how a ML model will behave in a certain decisional framework. As these solutions are not able to provide sufficient information about the model strengths and weaknesses, they fail to enable users to foresee when prediction errors might occur.[63] Addressing this problem, a research branch named 'third wave AI' aims to design systems with embedded explanation capacity that allow them to characterize real world phenomena. Also known as explainable-AI (XAI), the approach challenges certain problems also partially addressed in this article: i) classification or representation of events, objects or persons in heterogeneous data environments, and ii) constructing decision policies for autonomous systems.[64]

It is important to note that ML and the sphere of automated decisions are not monolithic and they have bifurcated implications resulting with diversely harmful effects. Rule-based modelling deals with the type of harms which may be contested on nor-

55 This transparency-gaming effect is known as Goodhart's law 'when a measure becomes a target, it ceases to be a good measure' Tal Zarsky, 'Transparent Predictions' (2013) Univ of Ill L Rev 4.

56 Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke L Tech Rev 18.

57 Leese (n 9) 505-506.

58 Joshua A Kroll et al 'Accountable Algorithms' (2016) 165 U Pa L Rev 633, 679.

59 LIME (Local Interpretable Model-Agnostic Explanations) is a 'model induction' technique that experiments with any given machine learning model—as a black box—to infer an approximate, an explainable model.

60 Marco Tulio Ribeiro, Sameer Singh and Carlos Guestrin, 'Why should I trust you?: Explaining the predictions of any classifier' (2016) Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 13–17 August 2016; Daniel Kasenberg, Thomas Arnold and Matthias Scheutz, 'Norms, Rewards, and the Intentional Stance: Comparing Machine Learning Approaches to Ethical Training' (2018) Proceedings of the Thirty- Second AAAI Conference on Artificial Intelligence <https://hrilab.tufts.edu/publications/kasenbergetal18aies.pdf> accessed 29 November 2018.

61 Marko Robnik-Šikonja and Igor Kononenko, 'Explaining Classifications for Individual Instances' (2008) 20 IEEE Transactions on Knowledge and Data Engineering 589 <http://lkm.fri.uni-lj.si/rmarko/papers/RobnikSikonjaKononenko08-TKDE.pdf> accessed 29 November 2018.

62 David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn about Machine Learning' (2017) 51 UC Davis Law Review 653.

63 Wenbo Guo et al, 'Towards Interrogating Discriminative Machine Learning Models' (2017) arXiv:1705.08564.

64 David Gunning, 'Explainable Artificial Intelligence (XAI)' (Defense Advanced Research Projects Agency, DARPA/I2O) <https://www.darpa.mil/program/explainable-artificial-intelligence> accessed 29 November 2018.

mative grounds such as unfair treatment or due process violations, primarily arising in *supervised ML*. There exist other methods for detecting and ameliorating various types of harms—eg, invasiveness, group-level harms[65], harms from economic manipulation and exclusion, and last but not the least, harms at the social and political level such as impairment of human dignity and self-autonomy—which may be hard to tackle in an individual contestation scheme.

V. Conclusion: Preliminaries of a Contestation Scheme

Rather than reflecting the underlying computational processes, RbM reverse engineers the result for a reconstruction of the decision-making process. By doing so, we employ a 'synthetic' approach aiming to acquire an understanding of the automated decision-making systems by means of model-building. The potential value of this effort is not merely a translation of the behaviour of ML systems in if-then statements. The distinction made between different types of normativity helps construct a bridge to 'causality' to expose and understand how the decision relates to our existence in this world. Whatever may be the chain of causation or the sophisticated(ness) of the inferences, as we regress far back, we will ultimately discover some input (information, statement, etc) which is conditioned by a certain representation as to the state of the world.

As an approach that intends to capture the concept of transparency by 'saying' what is required for the possession of it, the idea here is not to interpret a foreign domain through legal knowledge but to define requirements which would render data-driven systems more responsive, communicative and engageable from the legal or regulatory perspective. The proposed model should not be seen as a top-down initiative ordering system owners and engineers how they should design their systems. It is rather a bottom-up call from the view of the informed citizen simply formulating what the totality of the data-driven activities entail for review and contestation. It may be regarded as an overarching framework both intended as a guidance for the design and the audit of the automated decision-making systems, and also as a scheme for the *ex-post* scrutiny of specific decisions on several grounds and against different actors.

Based on the findings of this article, a contestation scheme, for instance as required by Article 22 of the GDPR, will involve a cumulative evaluation of the below questions:

- Is the (training) data representative of the data subject(s)? To what extent do the discrepancies matter—considering the purposes and the further impact of the decision as well as the regulatory context?
- Do the data features, selected and transformed, reflect (sufficiently construct) the reality (phenomenon under observation—eg the data subject) in a suitable, reliable and verifiable way for the given context and purposes?
- Based on these representations and constructs (decisional input) are the consequences 'explainable' by providing legally, ethically and socially agreeable reasons?
- Are the results interpreted and implemented as declared by the operators and designers of the system—the purpose of processing— and as mandated by law (purpose limitation)?
- Do the data subjects know to whom they should appeal and eventually hold accountable?

Where those responsible fail to respond to these contestability requirements, their automated decisions may be regarded as *per se* unlawful[66], or as ethically questionable, depending on whether or not any legal norms are violated. According to this view, intelligent systems may *only* be used if their underlying reasoning can be (adequately) explained to the targeted individuals.[67]

As mentioned in the previous section, in many cases technical transparency (disclosures) may be limited mostly due to the competitive or integrity-related concerns of the system operators (data controllers). ML-based models which fail to comply with certain contestability requirements may only be permitted under exceptional conditions together with strict scrutiny and pre-registration measures.[68] On the oth-

65 Linnet Taylor, Luciano Floridi and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Springer 2017).

66 Hildebrandt (n 25) 58.

67 Malin Eiband, Hanna Schneider and Daniel Buschek, 'Normative vs Pragmatic: Two Perspectives on the Design of Explanations in Intelligent Systems' ExSS '18, Tokyo, Japan, 11 March 2018.

68 Mireille Hildebrandt, 'Preregistration of machine learning research design. Against P-hacking' in Bayamlıoğlu et al (n 15); Margaret Mitchell et al, 'Model Cards for Model Reporting' (2018) <https://arxiv.org/abs/1810.03993> accessed 29 November 2018.

er hand, a satisfactory standard of contestability will be imperative in case of threat to individual dignity and fundamental rights.[69] It is also argued that the 'human element' of judgment is, at least for some types of decisions, an irreducible aspect of legitimacy in that reviewability and contestability are seen as concomitant of the rule of law and thus, crucial prerequisites of democratic governance.[70] In a wider perspective, we are confronted with the question whether there are functions, decisions and roles which should be confined to humans at all cost and all times. 'Is there any essence of humanity which should not be transferred to machines under any circumstances?'[71]

69 Alessandro Mantelero, 'AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment' (2018) 34(4) Computer Law & Security Review 754-772.

70 Zerilli et al (n 36); Emre Bayamlıoğlu and Ronald Leenes, 'The "rule of law" implications of data-driven decision- making: A techno-regulatory perspective' (2018) 10(2) Law, Innovation and Technology 1 <https://doi.org/10.1080/17579961.2018.1527475 >.

71 Thomas Burri, 'Machine Learning And The Law: 5 Theses' [2017] SSRN Electronic Journal.

Chapter 4 (*Third paper*)

The right to contest automated decisions under the GDPR: Beyond the so-called right to explanation

• Emre Bayamlıoğlu, "The right to contest automated decisions under the GDPR: Beyond the so-called right to explanation" (2021), *Regulation & Governance*; Vol. 15 Special Issue*: Algorithmic Regulation* eds. by Karen Yeung and Lena Ulbrich.

Regulation & Governance

Regulation & Governance (2021) doi:10.1111/rego.12391

The right to contest automated decisions under the General Data Protection Regulation: Beyond the so-called "right to explanation"

Emre Bayamlıoğlu

KU Leuven Centre for IT & IP Law (CITIP), Leuven, Belgium

Abstract

The right to contest automated decisions as provided by Article 22 of the General Data Protection Regulation (GDPR) is a due process provision with concrete transparency implications. Based on this, the paper in hand aims, *first*, to provide an interpretation of Art 22 and the right to contest (as the key provision in determining the contours of transparency in relation to automated decisions under the GDPR); *second*, to provide a systematic account of possible administrative, procedural, and technical mechanisms (transparency measures) that could be deployed for the purpose contesting automated decisions; and *third*, to examine the compatibility of these mechanisms with the GDPR. Following the introduction, *Part II* starts with an analysis of the newly enacted right to contest solely automated decisions as provided under Article 22 of the GDPR. This part identifies the right to contest in Article 22 as the core remedy, with inherent transparency requirements which are foundational for due process. Setting the right to contest as the backbone of protection against the adverse effects of solely automated decisions, *Part III* focuses on certain key points and provisions under the GDPR, which are described as the *1st layer* (human-intelligible) transparency. This part explores to what extent "information and access" rights (Articles 13, 14, and 15) could satisfy the transparency requirements for the purposes of contestation as explained in Part II. Next, *Part IV* briefly identifies the limits of *1st layer* transparency – explaining how technical complexity together with competition and integrity-related concerns render human-level transparency either infeasible or legally impossible. In what follows, *Part V* conceptualizes a *2nd layer* of transparency which consists of further administrative, procedural, and technical measures (i.e., *design choices facilitating interpretability*, *institutional oversight*, and *algorithmic scrutiny)*. Finally, *Part VI* identifies four regulatory options, combining *1st and 2nd layer* transparency measures to implement Article 22. The primary aim of the paper is to provide a systematic interpretation of Article 22 and examine how "the right to contest solely automated decisions" could help give meaning to the overall transparency provisions of the GDPR. With a view to transcend the current debates about the existence of a so-called right to an explanation, the paper develops an interdisciplinary approach, focusing on the specific transparency implications of the "right to contest" as a remedy of procedural nature.

Keywords: algorithmic transparency, algorithmic regulation, automated decisions, GDPR.

> *[…]and there is some temptation to obey the computer. After all, if you follow the computer you are a little less responsible than if you made up your own mind. (Bateson 1987, p. 482)*

1. Introduction and outline

Increasing automation has been an important topic of concern even at the earliest stage of the debates about the legal, political, and economic impact of data practices in the digital realm. It was clear by the early 1970s that resentment engendered by the systems such as *computerized billing* would soon spill over onto more delicate domains of life. Cautions were expressed that automated data processing would impair the system operators' capacity to provide explanations about the results produced by the system and thus, contribute to the "dehumanizing" image of computerization.

Correspondence: Emre Bayamlıoğlu, Sint-Michielsstraat 6, Box 3443, 3000 Leuven, Belgium. Email: emre.bayamlioglu@kuleuven.be

Conflict of interest: The author declares that there exists no conflict interest and no relevant data to be made available.

Accepted for publication 6 February 2021.

Based on these concerns, the notion of transparency has long been regarded as a means to limit the risks and mitigate the harms arising from the opaque nature of data processing. Since the enactment of the Data Protection Directive (DPD) in 1996, the foundational idea underlying the EU data protection regime has been that the adverse effects of data processing may be best addressed by permitting individuals to learn about the data operations concerning them. Today, with the General Data Protection Regulation (GDPR), the European data protection regime may now be considered as the most extensive body of law aiming to regulate the activities involving personal data. It not only maintains well-defined individual rights fleshing out the principle of transparency but also accommodates various tools and mechanisms for the implementation and enforcement of these rights.

With data-driven practices based on machine learning (ML) being the primary *foci* of the data protection reform which resulted in the GDPR, one of the novelties of the Regulation is the enhanced transparency scheme provided for solely automated decisions – in particular, the introduction, the right to human intervention, and right to contest in Article 22.[1] Accordingly, the paper in hand deals with this specific type of transparency, namely "transparency" in the sense of *interpretability* for the purpose of contesting automated decisions. The aim is to determine to what extent the GDPR accommodates the practical implications of "right to contest" and the ensuing transparency requirements.

Taking right to contest as a due process provision, Part II starts with a systematic interpretation of Article 22, examining how the concepts of *contestation*, *obtaining human intervention*, and *expressing one's view* should be understood and interrelated. Rather than a prolongation of the initial provision (Article 15 of the DPD), the right to contest is regarded as the backbone provision with a key role in determining the scope of algorithmic transparency under the GDPR. To fully lay out the transparency implications of the right to contest, this Part also addresses the question: *what should be made transparent or known in order to render automated decisions interpretable and thus contestable on a normative basis?* (Bayamlıoğlu 2018). The analysis inquires what interpreting the "algorithm" could mean for the purpose of contesting automated decisions – confirming that the transparency implications of right to contest are too complex to be dealt with merely by addressing certain opacities or invisibilities.

Overall, Part II lays out the theoretical basis of the paper approaching *data processing* and *automated decision-making* (ADM) as regulatory technologies, which enable a form of "algorithmic regulation"(Yeung 2018).[2] Such techno-regulatory approach allows for a conceptualization of ADM and the surrounding transparency debate as *a procedural*, put in other words, as a *due process problem*.[3] Therefore, instead of handling automated decisions from the narrow lens of *discrimination*, *bias*, or un*fairness*, this paper regards ADM systems as "procedural mechanisms" which produce legally challengeable consequences. The concepts like fairness, equality, or nondiscrimination – as being mainly contextual and domain-dependent – can only address a fragment of the problem and thus, cannot serve as a theoretical basis for the intended analysis. Moreover, misuse of these quasi-legal concepts (to give meaning to the statistical results) runs the risk of *technical solutionism*, which will misinform policy-makers about the ease of incorporating transparency and accountability *desiderata* into ML-based systems (Cath 2018, p. 3, 4).

Having laid out the transparency implications of Article 22 as a general provision of due process, Part III, IV, and V inquires to what extent the GDPR can accommodate different conceptions of transparency inherent in the right to contest.[4] The analysis is based on a twofold approach. That is, the "information and access" rights (Article 13–15) and the safeguards (Article 22) are treated as complementary but distinct sets of remedies (as *1st* and *2nd layer* transparency).[5] This twofold methodology is guided by the understanding that recognizing distinct forms of opacity inherent to ADM systems is vital in developing (technical and nontechnical) solutions to address the risks arising due to the impenetrable nature of ML (Burrell 2016, p. 2).

In what follows, Part III focuses on certain key principles and provisions under the GDPR, which we describe as the *1st layer* (human-intelligible) transparency. It explores to what extent "information and access" rights in Articles 13, 14, and 15 of the GDPR could facilitate or improve the contestability of automated decisions as explained in Part II.

Part IV briefly identifies the limits of *1st layer* transparency, explaining how technical constraints together with the competition and integrity-related concerns (of the system developers and operators) render human-level transparency infeasible or legally impossible. Reflecting on both technical and economic limits, this Part offers an account of why the transparency requirements for contesting automated decisions could not be limited to access, notification, or explanation in the conventional sense.

Having seen the limits of directly human-intelligible models based on disclosure and openness in the previous Part, Part V inquires what further solutions the GDPR could accommodate in terms of implementing different conceptions of transparency aiming for contestability. As the *2nd layer transparency*, this part systemizes various regulatory instruments and techniques under a threefold structure: (i) the design choices facilitating interpretability; (ii) the procedural and administrative measures; and (iii) the software-based tools for algorithmic scrutiny.

Given that the problem lays with the framing of the optimum extent of transparency and the appropriate mode of implementation, Part VI offers regulatory options (implementation modalities) combining *1st and 2nd layer* transparency with a view to implement Article 22 without prejudice to the integrity of the systems or the legitimate interests of the stakeholders.

The final Part concludes that despite the normative, organizational, and technical affordances explained throughout the paper, between the right to contest as provided in the GDPR and its practical application, there are many gaps to be bridged to achieve the desired level of protection without hindering data-driven businesses and services. Accordingly, the conclusion points out the relevant research domains where further progress is required to construct a compliance scheme capable of balancing competing interests. Hence, the paper also serves as a conceptual framework for future research aiming to unravel sector or domain-specific barriers in relation to implementation of the right to contest,

With a view to transcend the current debates surrounding the so-called right to an explanation, the paper conceptualizes ADM as a regulatory technology and focuses on the specific transparency implications of the "right to contest" as a remedy of procedural nature. Building on the former writings of the author about the transparency implications of ADM (Bayamlıoğlu 2018), the main contribution of the paper lies in this procedural perspective – which enables an interpretation of Article 22 as a due process provision – followed by a systematic analysis of the possible implementation tools and modalities under the GDPR.

2. Article 22 of the GDPR and the right to contest automated decisions

The principle laid out in Article 22, requiring that the automated data-driven assessments cannot be the sole basis of the decisions about the data subjects, is unique to the EU data protection regime. Such provision is not generally included among the US fair information practices or in the OECD guidelines preceding the 1996 DPD (Edwards & Veale 2017). Article 22 does not directly target personal data processing but a certain type of outcome, that is, the decisions that are fully automated and that substantially affect individuals.[6]

Since the enactment of the DPD in 1996, the practical application and proper implementation of Article 15 (the precursor to Article 22 of the GDPR) has not been of concern neither to the supervisory nor to the judicial authorities (Korff 2010). Although the provision was found intriguing and forward-looking, due to its complex nature – which makes individual enforcement difficult – it has been mostly overlooked and underused (Mendoza & Bygrave 2017). In practice, the compliance standards of the provision have remained at *de minimis* level, reducing compliance to a mere formality. According to Zarsky: a rule which is rarely applied (Zarsky 2017, p. 1016).

At first glance, Article 22 of the GDPR may be seen not to have brought much change in terms of wording. In this regard, the initial formulation of the provision in 1996 seems to have made it somehow future-proof. However, as will be explained below, with the newly introduced safeguards (the right to human intervention and contestation), the provision now has an essential role in determining the scope of transparency for solely automated decisions under the GDPR.

2.1. Decisions based solely on automated processing, with legal or similarly significant effects

The key provision of the GDPR on ADM, Article 22, applies to processes, which are fully automated, and which bring about legal or similarly significant effects for the data subject. Automated decisions, which fail to comply with the definition provided in Article 22(1), shall not be bound with the provision.

The application of Article 22 initially requires the existence of a "decision," though neither the former DPD nor the GDPR provides any guidance as to what amounts to a decision. Bygrave suggests that the term "decision"

should include similar concepts such as *plans*, *suggestions*, *proposals*, *advice*, or *mapping of options*, which somehow have an effect on its maker such that she/he is likely to act upon it (Bygrave 2001).

Article 22(1) further requires that the decision must also be fully automated, allegedly involving no human engagement. Because the level of human intervention to render the decision not fully automated is not clarified, many data controllers interpret the provision narrowly. As a result, significant amount of data-driven practices may be kept out of the reach of EU data protection regime simply by the nominal involvement of a human in the decision-making process.[7] This requirement which also existed in the DPD has been widely used as a loophole by the data controllers to derogate from the provisions on automated decisions. This has been despite the preparatory work of the DPD, which explicitly stated that one of the rationales behind Article 15 of the DPD was that human decision-makers might attach too much weight to the seemingly objective and incontrovertible character of sophisticated decision-making software – abdicating their own responsibilities.[8]

The scope of Article 22 is limited to the decisions that produce legal or similarly significant effects. Legal effects may be described as all qualifications established by a legal norm either in the form of obligations, permissions, rights, powers; or in relation to one's status such as citizen, parent, spouse, debtor; or relating to categories of things (e.g. moveable, negotiable instrument, public domain). The inclusion of the term *similarly significant effects*, expands the scope of the provision to cover certain adverse decisions even if the outcome does not straightforwardly affect the data subjects' legal status or rights.

Regarding the implementation of Article 22 in the EU, some member states have adopted a wider approach, such as Hungary, which includes all automated decisions prejudicial to the data subject or France, where the specific legislation covers ADM producing any significant effect (Malgieri 2019).

2.2. Derogations: Consent, contractual necessity, and mandatory laws

As Article 22(1) grants data subjects the right not to be subject to solely ADM, the provision also contains certain exceptions (derogations) to the rule – subject to Art. 22/4 on special categories of personal data.

One of the most important changes brought by the GDPR as compared to Article 15 of the DPD is the introduction of "*explicit consent"* in Article 22(2)(c) as one of the grounds, which may be relied upon by the controllers to carry out fully automated decisions. According to Mendoza and Bygrave (2017), the introduction of consent comes as an impairment to the essence of the provision, lowering the *de facto* level of protection (p. 96). Considering that *consent may practically be used to deprive the data subjects of the control of their data*, the concerns about this new derogation – as a swift mechanism to carry on with automated decisions – are not all without merit. Nonetheless, rather than serving as a backdoor to circumvent data protection rules, consent may equally be construed as a leverage for transparency (Kaminski 2019). This is particularly the case where explicit and informed consent is taken as the initial step of the safeguards to render automated decisions contestable under Article 22(3). Furthermore, consent does not relieve the data controller from the duty of compliance with the general data protection principles such as fairness and proportionality provided in Article 6. Taking into account the complexity and subtlety of the current ADM systems, the requirement of explicit consent inevitably entails some "explanation" to allow data subjects to make informed choices.[9] The extent of communication necessary to render the data subject's consent explicit and thus valid, may also be taken as a benchmark to determine the minimum content of the notifications under Articles 13 and 14. Consent implemented as a leverage to individualized transparency – but not as a *carte blanche* for ADM without encumbrance – could play a critical role in reinforcing the right to obtain human intervention and right to contest.

Formation and performance of a contract (contractual necessity) is another derogation provided in Article 22(2)(a). The prohibition on automated decisions shall not be applicable where an automated decision is necessary for entering into or for the performance of a contract between the data subject and the data controller. The derogation based on contractual necessity provides a broad field of play which is tempting for abuse and *creative compliance*. The extent of automated decisions, which would be necessary in a contractual context is an issue that requires the consideration of the mutual benefits and expectations of the parties. For instance, increasing the efficiency of the system – as a general argument – cannot be regarded as necessity for this is simply what makes data processing more invasive (Guinchard 2017, p. 12).

Article 22/(2)(b) lays out another derogation providing that data subjects may be deprived of the safeguards in Article 22(3) where processing is mandated by the Union or Member State Law. Despite the reference to *suitable measures to safeguard the data subject's rights and freedoms and legitimate interest*s, this exclusion is likely to create discrepancies in terms of contestability standards between the administrative decisions based on public law and ADM relying upon consent or contractual necessity (Massé & Lemoine 2019, p. 9). Confirming this, some member states have already implemented the derogation in a way similar to a blanket-exemption, permitting ADM as default practice for public institutions (Malgieri 2019).

2.3. Safeguards against automated decisions and the right to contest

2.3.1. Safeguards in Article 22(3) in general

Under Article 22(3), where the exemptions based on contractual necessity (22 (2)(a)) or consent (22 (2)(c) take effect, the data controller is obliged to implement measures to safeguard the data subject's rights, freedoms, and legitimate interests. In principle, these measures should at minimum contain a fair amount of human intervention so that the data subjects may express their view and effectively contest automated decisions. Before the GDPR, the DPD only spoke of arrangements allowing the data subjects to put forward their point of view. The Regulation has improved this position by formulating safeguards providing *for human intervention* and *contestation*.

Article 22(3) reads as:

> *In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.*

Although frequently commented and explored in the scholarly writing, there seems not to be much attention to a coherent and systematic interpretation of the provision, in particular how the right to obtain *human intervention*, *expressing one's views*, and *contesting* the decision could practically be implemented. They are usually treated as if rights or remedies of equal footing (alternatives to each other), without clarity whether they are complementary, gradual, or distinct rights, or they should be treated as a unity.[10] Very few seem to pay attention to the inevitable difficulties arising from the provision's open-ended and clumsy syntax – dressed in dense and contorted "legalese" – muddying its interpretation (Yeung & Bygrave 2020, p. 10). Wachter and others rightly point out that whether these remedies are "interpreted as a unit that must be invoked together, or as individual rights that can be invoked separately, or in any possible combination" would determine how a decision could be contested. After acknowledging the possibility of several interpretations, they conclude that treating each Article 22 safeguard as "individually enforceable" would be the most sensible option (Wachter *et al.* 2017).

More importantly, Wachter et al. (2017) also argue that – depending on the costs and the "likelihood of success" – expressing views or obtaining human intervention should not necessarily be followed by further legal means to challenge the decision. However, a systematic and teleological reading of the provision reveals that the right to contest is the backbone of the safeguards provided under the GDPR and cannot be ignored on the mere grounds of low likelihood of success as Wachter *et al.* suggest. The wording makes it clear that the right to obtain human intervention is the minimum of the remedies that data controllers are obliged to implement to satisfy the ultimate aim of the provision. This necessarily implies that human intervention may not always be the best option to address the adverse effects of automated decisions. As Part V will explain, there may be further options (other than or in addition to human intervention), which provide technical and procedural means to challenge the ML outcome. Specific inclusion of the "right to express one's point of view" confirms that data subjects not only have a right of appeal to obtain a new decision but they can also provide information that might be relevant for reconsidering the initial result (Malgieri 2019, p. 11). There may also be dignitary concerns or possible mutual benefits in allowing the data subjects to express their views even if they are not willing to challenge the decision. For instance, through an online forum, data subjects can provide feedback and voice their complaints about the extenuating circumstances that they believe the system fails to consider.

Due to its procedural character, Article 22/3 is inevitably silent on the substantial grounds which could be relied upon to challenge the reasoning or the criteria underlying the automated decisions (*see* Part 2.4). That is,

whether or when certain ML outcome could be regarded as unfair or unlawful is a conclusion, which requires resorting to normative propositions provided in the relevant legal domains, for example, labor law, consumer law, insurance law (Macmillan 2018, p. 51). Data protection law is not where we could seek for standards or rules that can be used to test the legality or the justifiability of the undesirable outcomes of ADM (Hoffmann-Riem 2020, p. 14).[11]

2.3.2. The right to contest

The newly adopted wording of the GDPR using the term "contest" connotes more than mere opposing. It rather points at a "right of recourse" as a version of algorithmic due process (Kaminski 2019a) or, at least, an obligation to hear the merits of the appeal and to provide a justification for the decision. Although the essence of the provision has not changed with the introduction of "human intervention," the inclusion of an appeal (contestation) process against automated decisions stretches the legal boundaries of Article 22 to the broadest extent possible. It obliges the data controller either to render automated decisions contestable or to cease ADM at all. What is required by Article 22(3) is not about informing or disclosing but rendering the decision contestable at least against a human arbiter. In principle, the data controller has to "explain" the decision in such a way that enables the data subject to assess whether the reasons that led to a particular outcome were legitimate and lawful (Goodman & Flaxman). Confirming this, German law – long before the GDPR – has taken the view that the data subject needs to have an understanding of a specific decision in order to evaluate and be persuaded about its accuracy and legal compatibility (Korff 2010, p. 83).

Borrowing from the Contract Law of civilian (continental) legal systems, the distinction between the access and information rights (Art. 13–15 GDPR), and the due process rights in Article 22(3) may be seen similar to that of "obligations of conduct" and "obligations of result," respectively. An *obligation of conduct* requires the dedication of an adequate amount of resources or the use of reasonable endeavors for a certain end, without any guarantees as to the outcome (a.k.a. input-based obligation) (Economides 2010). From this perspective, duties pertinent to access rights may be found more akin to an obligation of conduct since "informing," "explaining," or "disclosing" are acts which refer to a certain behavior, rather than assuring a specific outcome. The right to contest under Article 22, on the other hand, is similar to an *obligation of result*. It mandates a mechanism that will enable data subjects to have their objections heard and decided. This theoretical distinction helps draw a coherent picture of the contestation scheme provided by the GDPR for automated decisions. That is, contestability is a goal-oriented concept, unless the desired outcome is achieved, what is disclosed or communicated is irrelevant. In line with this, Kroll rejects the view that computer system behaviors should be held to a "best-effort" standard because they are unforeseeable. He asserts that treating these systems as though they are uncontrollable would ignore the fact that they are human artifacts – built to a purpose by some human agency that must be accountable for their behaviors (Kroll 2018, p. 7).

Despite this central role of the right to contest in Article 22(3) – which somehow presupposes a form of explanation of the decision – some scholars, nevertheless, argue that the GDPR does not contain an *ex-post* "explanation right" for individual automated decisions (Wachter *et al.* 2017). To support their argument, the proponents point out the discrepancy between the wordings of Article 22(3) and Recital 71. That is, while the latter specifically mentions obtaining an "explanation of the decision," the former does not. Accordingly, it is contended that the GDPR provided no right of explanation for individual automated decisions (Wachter *et al.* 2017). Leaving aside the objections raised to this rigid and contained interpretation, the approach taken in this paper renders such argument and the surrounding debate mostly irrelevant. It is inherent in the formulation of the right to obtain *human intervention* and *contestation* that Article 22(3) requires much more than a mere explanation of the decision. As will be explained in Part III below, access and information rights relating to automated decisions can only be effectively enforced if they contribute to the due process rights provided in Article 22. The omission of the phrase "an explanation of the decision" from Article 22(3) may be understood as the intention of the legislature to keep an open view as to the safeguards and their possible implementation. Considering that the *travaux préparatoires* of the Regulation do not provide much clarity about the rationale of the provisions relating to automated decisions, the legislature seems to acknowledge that such novel legal obligation, expressed in abstract terms, could not be properly contained or effectively addressed by the concept of "explanation." Therefore, the provision may be regarded as giving leeway for different methods of implementing safeguards. Article 22 of the

GDPR not only subsumes the much-discussed "right to explanation" but also allows for different regulatory options or modalities to ultimately provide the data subjects with the means to contest automated decisions. As Selbst and Barocas (2017) put it: "…focusing on explanation as an end in itself, rather than a means to a particular end, critics risk demanding the wrong thing" (p. 1).

These above-explained controversies and ambiguities are also reflected in the implementation of Article 22(3) within the EU. Member states formulate provisions on ADM in various ways without much coherence, and few expressly refer to explanation. French Law, as providing the most generous framework, permits data subjects to obtain information about the rules of the data processing and the features relating to the practical implementation of the algorithm together with the source code. Hungarian law, with a rather innovative formulation, requires information *about the methods and criteria used in the decision-making mechanism* (Malgieri 2019).

2.4. The transparency implications of the right to contest (contestability requirements)

Having explored the possible interpretation of the right to contest as provided in Article 22, this Part – building on the author's earlier work – provides a brief account of the possible transparency(contestability) requirements to effectively challenge automated decisions. The below analysis reflects on a dynamic and instrumental conception of transparency, which do not aim to analyze the system by the semantic route of explanation but rather by defining requirements enabling scrutiny on a normative basis (Bayamlıoğlu 2018; Rader *et al.* 2018). Put in other words, contesting ML-based decisions is not about reading off the computer code but rather relates to the question of how these systems make up the regulatory realm we are subjected to.

Since machines are built for a purpose, they are expected to exhibit certain behaviors associated with their function (De Ridder 2006). That is, every decision-making system contains some inherent "normativity"[12] as the system's output is directed to achieve some preset goals or to serve certain ends (Castañeda 1970; Krist 2006; Binns 2017). As mentioned above, ADM systems may also be seen as techno-regulatory assemblages, which select and reinforce certain values at the expense of others (Bayamlıoğlu & Leenes 2018; Eyert *et al.* 2021). Accordingly, challenging an automated decision initially requires a conceptualization of the outcome as a process where certain input leads to certain results. This may be seen akin to the decisions in a legal system, based on "facts," "norms," and the ensuing "legal effects." Such conceptualization aligns with the approach, which portrays regulation as a cybernetic process involving three core components that form the basis of a control system, that is, ways of gathering information, setting standards, and ways of changing behavior (Yeung 2015). In the context of automated decisions, this would imply how and why a person is classified/profiled in a certain way, and what consequences would follow from that classification. Such modeling, which maps input/data with the effects within a contemplated normativity, provides us with a rule-based reconstruction of the decision-making process. (Bayamlıoğlu 2018). The contestation of a decision relates both to the interpretation of the input and the normative basis relied upon to reach that decision (Hoepman 2018). How transparency is effected will depend on what it is intended to accomplish. The concrete transparency requirements of such a model entail numerous types of differently purposed, but complementary, information flows along multiple axes (Malgieri & Comandé 2017; Kaminski 2019a). In the below paragraphs, we briefly provide the essentials of a model, which aims to open up and systemize what interpreting the "algorithm" could mean for the purpose of contesting automated decisions.[13]

For computers to solve a problem, it is necessary that a computation manipulates a representation of the world, and the meaning of a computation depends on the meaning of the representation it transforms. Therefore, what ML systems do may be regarded as the creation of *internal models* of the pertinent environments (Eyert *et al.* 2021). Based on this, as the *initial* step of contestation, we need the knowledge of *what the system learns about persons*, *places or events*, and *how people are represented as inputs to the algorithm*. In a ML process, data instances exist as values of feature variables where each feature (e.g. age, height, weight) is an individually measurable dimension of the problem in question. Determining which data features to consider is a part of the regulatory process as it reduces the complexity of the environment to a specific segment of "reality" (Eyert *et al.* 2021). As the necessary interface between the underlying attributes and the decisions that depend on them, data features are open normative challenges. Accordingly, decisions may be contested based on the selection of the relevant data features and the ensuing inferences that are relied upon. Or as Jasanoff and Simmet (2017) put it: "…the choice of which realities one takes as consequential and therefore which facts one sees as important or

controlling, is normative" (p. 752).[14] Having said that, it should be noted, ML-based systems are less and less programmed with a predefined feature space. Deep learning techniques using neural networks can define features autonomously by analyzing the data directly coming from the input layer. This severely impedes the capacity to scrutinize the factual or inferential basis of any decision or outcome.

When ML tools are used to make decisions, it is possible to contemplate a "decision rule" such as: *do this if the estimated probability of "z" is larger than "x" or smaller than "y," and so on* (Baer 2019, p. 88). Thus, the *second* type of insights required for contestation is the decision rules (normative basis of the decision), which describe how certain ML findings are translated into concrete results in a wider decision-making context. For instance, speech analysis – which can detect one's dialect or accent – may be used as "factual" input for the selection of the suitable content in political microtargeting. Accordingly, Spanish voters identified as having a Catalan accent may be delivered political messages supporting Catalunya's secession from Spain. In this example, the delivery of the relevant content is the result of applying "decision rules" to the outcome of the classifier. Decision rules (normative choices) are shaped by the hypotheses and assumptions about the root cause of the targeted problem. They are the formalizations of the general goals (objectives) of the system such as winning elections, avoiding customer churn, or better distribution of insurance risks. ML-based systems allow for various combinations of general goals and the ensuing decisional rules "nested inside one another" (Eyert *et al.* 2021). Since this normativity does not necessarily rely upon legal or moral grounds but has a computational and data-driven basis, the normative orientations of the system are not always as straightforward as the relation between having Catalan accent and supporting independence. It is often the case in ML practices that the objectives and the underlying assumptions may not be deterministically configured but rather adaptively and dynamically adjusted (Yeung 2018).

Third, the "impact" of the decision could also be an essential ground for contestation. Speaking of the impact of the decision, a simple credit score does not only determine whether one would get a loan, but it may also, fully or partially, determine the loan pricing, type of loan monitoring, the amount of credit, and how the credit risk would be managed. As such, for the purposes of contestation, the "context" determines the actual consequences (impact) of the decision. A decision may be "good" in a particular context but less "good" in others (Zeng *et al.* 2018). In case of contextual uncertainty, there could be several explanations which may seem equally plausible (Gollnick 2018).

Fourth, information about the "accountable actors" behind the ADM process is also an essential piece of information necessary for effective contestation. Although, this may seem like a procedural requirement – not primarily of relevance to interpreting the decision in substance – it is nevertheless vital to fully understand the context and the purposes underlying the decision-making process. In case of automated decisions, there may be several legal grounds of contestation relating to different actors (e.g. errors in data collection, flaws in analysis, illegitimacy of the purposes, or the ensuing decision rules). As such, challenging an automated decision not only requires the disclosure of the entities involved but also the organizational and contractual set-up under which these entities operate and conduct business. The GDPR contains various provisions, which may accommodate disclosures reaching beyond the conventional actors of "data controller" and "data possessor." The reference made to "the recipients or categories of recipient of personal data" in Articles 13, 14, and 15, and further definitions provided for "representatives," "group of undertakings," and "enterprises" are clear indications that the GDPR was drafted in consideration of the complex network of actors behind the current data-driven practices.

3. A general overview of "Access and Information Rights" (*1st layer transparency*)

This part will explore how the individual information and access rights of the GDPR (*1st layer* of transparency) could substantially contribute to the objectives of Article 22. Put in other words, the below analysis explores: to what extent the relevant provisions in Articles 13 to 15 could be interpreted in the direction of "contestability" as defined above.

3.1. The intended purposes and the legal basis of processing

Articles 13(1)(c) and 14(1)(c) of the GDPR provide that data subjects will be given information about the purposes of the processing for which the personal data are intended together with the legal basis for processing. As a

reflection of the principle of purpose limitation (purpose specification) in Article 5(1)(b), information about the "intended purposes" is a key element, which helps reveal the business strategy and the objectives pursued by the data controller as well as the other related parties. Such information enables the "reverse-engineering" of the decision-making process with a view to understand the underlying normative setup. The purposes pursued by the system are also of direct relevance to the context of the decision.

In case the data controller relies on Article 6(1)(f), which lays the general grounds for lawful processing, the obligation to notify the data subject of the intended purposes is further reinforced by the requirement to provide *information about the legitimate interests pursued by the controller* (Articles 13(1)(d) and 14(2)(b)). In addition, the reference made to the *legal basis of processing* (statutory or contractual) in Articles 13(1(c) and 14(1)(c) makes it clear that the GDPR envisages a link between the *intended purposes*, *pursued interests*, and the *legal basis* of data processing. These altogether may be implemented as effective transparency mandates against the data controllers that carry out solely automated decisions. In this respect, Hildebrandt draws attention that the purpose of processing that is to be defined by the data controller is not the same as the tasks to be defined by the system designers to achieve that purpose. While the former is related to the commercial, institutional, political, or moral aims of those who deploy the system, the latter deals with the objectives and/or targets that the learning algorithms have been programmed to follow. Purpose limitation does not primarily aim for *the methodological integrity of data science*, but it is rather a specific reflection of the principles of legality and due process. As such, "it relates to the justification of such decision-making rather than its explanation in the sense of its heuristics." (Hildebrandt 2019, p. 113).

The principle of "purpose limitation and specification" also plays a key role in determining the extent of liability with regard to relevant stakeholders. In various cases, the CJEU has held that those who exert influence over the processing of personal data and participate in the determination of the purposes of that processing, may be regarded as a *controller* within the meaning of the EU data protection regime.[15] Accordingly, in *Fashion ID* case,[16] it was made clear by the Advocate General that the power to decide and specify for which purposes the data will be processed is a crucial factor in the apportionment liability among the involved parties.

Since the enactment of the DPD, data controllers have had trouble in deciding how to adequately specify the purpose in a certain data processing operation. So far, many data controllers have chosen to phrase their purposes as vague and abstract as possible. This is both to have the maximum leeway for further use of the data and also to avoid the disclosure of any commercially valuable information regarding their data operations. Taking that into account, in its 2018 *Guidelines on Automated individual decision-making and Profiling*[17] (hereafter *WP29 Guidelines on automated decisions*), Working Party 29 (WP29) made it clear that the purposes defined such as "improving users' experience," "IT-security," or "future research" would not suffice in the absence of further clarification. For instance, processing of data for online advertising may not be compatible if the initial notification only contained a mere reference to "marketing purposes."

In terms of exercising the right to contest, *purpose limitation*, *compatible use*, and *notification of the intended purposes* are important leverages, the implementation of which are dependent on certain enacted and applied transparency requirements.

3.2. "Meaningful information about the logic involved" and "the envisaged consequences"

Articles 13(2)(f), and 14(2)(g) GDPR provide that the controller shall inform the data subjects of: (i) the existence of ADM as defined in Article 22; (ii) meaningful information about the logic involved; and (iii) the significance and the envisaged consequences of the decisions. As seen, these provisions directly correspond to the essential constituents of the contestability requirements defined above, namely the "facts" (data input in the form of features) that are relied upon and the *decision rules* informing us about the goals pursued by the system.

As an initial step, the relevant provisions in Articles 13 and 14 require that the data controller provides information about the existence of a decision based on solely automated processes as defined in Article 22(1). Next, the data controller is obliged to provide *meaningful information about the logic involved* in the processing together with the *significance and the envisaged consequences* of the decision. Although emerging big data practices have been one of the major driving forces behind the GDPR, this crucial provision has remained similar to its DPD counterpart, which was mostly neglected or underused during the lifetime of the Directive. In more than

20 years that the DPD had been in force, the scope, requirements, and the possible limitations regarding the right of access as applied to automated decisions was not tested before the European courts. There is thus hardly any practical guidance on the interpretation of this enduring provision. So far, the obligation to provide information about the logic of the ADM has had varying implementations in the EU member states (Korff 2010, p. 85).

As the GDPR adds the term "meaningful" to the original provision in the DPD, it is now generally accepted that the controller should convey information about the rationale and the criteria relied upon in reaching the decision. The quality of being "meaningful" must be evaluated from the perspective of the data subject, treating accessibility and comprehensibility as the primary components. In parallel with the contestability requirements in Part II, "meaningful information" may also be understood as a functional description, which connects the decisional cues (data as input) with the consequences in a normative contemplation. Selbst and Barocas (2017) assert that "[t]he GDPR's demand for meaningful information requires either that systems be designed so that the algorithm is simple enough to understand, or can provide enough functional information about the logic of the system that it can be tested" (p. 31). Lipton (2016), drawing attention to Article 22, more elaborately states that the information to be conveyed must "(i) present clear reasoning based on falsifiable propositions and (ii) offer some natural way of contesting these propositions and modifying the decisions appropriately if they are falsified" (p. 4).

Further explicit reference in the provision to the "significance" and the "envisaged consequences" of processing resonates with the above-defined contestability requirement about the impact of the decision (Part 2.4). For the purposes of contestation, it is essential to fully understand the concrete results and the risks emanating from the contextual use of the data. For instance, in credit scoring, envisaged consequences may include whether the result of the analysis will be used for subsequent evaluations, the period during which the evaluation will be held valid or the third-parties who might have access to the results. The information about the "envisaged consequences" should elicit the real-life impact of the automated decisions to enable the data subjects to oversee the process and evaluate the consequences. The envisaged consequences should be assessed in tandem with the *intended purposes* of data processing.

In line with above, WP29 *Guidelines on automated decisions* clarify that "[t]he data subject will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis."(p. 27). The Guidelines recommend (to data controllers carrying out automated decisions) to provide information about "why a certain profile is relevant to the automated decision-making process and how the profile is used for a decision concerning the data subject" (Annex 1).

4. Limits of and impediments to human-interpretable models

The analysis thus far reveals that the GDPR provides several individual rights, which accommodate a fair amount of information to facilitate the exercise of the right to contest at a human-intelligible level (*1st layer transparency*). That being said, satisfying contestability requirements through openness, disclosure, and notification is neither desirable nor necessarily feasible for the purposes of challenging automated decisions (Desai & Kroll 2017, p. 39).

This Part further systemizes and examines where and why the *1st layer* transparency (information and access rights) fails – necessitating that the *2nd layer transparency*(Part V) come into play. The main impediments to *1st layer transparency* are: (i) the technical (complexity-related) intransparencies; and (ii) the secrecy demands of the businesses and institutions primarily arising from the integrity or competition-related concerns (Also see Mantelero 2019, p. 11).

4.1. The technical limits: Computational complexity and unpredictability

Human-level transparency (*1st layer*) for the purpose of contestation will mean that given enough time and resources, the computational processes producing the result should be human intelligible for the purposes of challenging the decision (Hildebrandt 2018). However, in practice, computational complexity gives rise to inscrutable models as a technical matter.[18] Exceedingly complex models are often very difficult or even impossible for humans to parse.

First, even in simple models, the rules that govern decision-making process may be so numerous and interdependent that they defy practical inspection and thus, resist comprehension (Selbst & Barocas 2018, p. 1094). Generally, the more factors are incorporated into the model as input, the more rules will be required to explain all possibly valid relations between the input and the output. In consequence, the system may end up with too many predictors, each having a weak relationship to the result.

A *second* type of opaqueness arises from the fact that the value of ML lies largely in its capacity to find patterns that go beyond human intuition. This results in ML models, which make it impossible to weave a sensible story to account for the statistical relationships that seem to weigh in. The assumed causality that the decision relies upon may be obscure and thus, may defy our intuitive expectations about the relevance of the criteria. Correlative relationships can be sufficiently complex and nonintuitive especially when dealing with human behavior. For instance, ML tools deployed for recruitment can decide about the best prospective employees according to the applicant's place of birth, music taste or, peculiarly, whether the applicant has any numerical characters in her social media account name. When the features used do not bear a comprehensible relationship to the outcome, the model will resist an assessment whether the decision is reliable – both as a matter of validity and as a normative matter (Selbst & Barocas 2018, pp. 1098, 1129).

Third, adaptive and dynamic data-driven systems are capable of modifying their responses according to the changes in the environment. Accordingly, in adaptive decision-making systems, the decision rule is no longer predetermined, but constantly adjusted (Yeung 2018). An adaptive or nondeterministic algorithm may produce different results for each instance of its execution (each case it handles). Therefore, while complexity can be seen as a barrier to overall understanding, adaptive algorithms seriously impair the capacity to predict the results for a particular set of input (Felten 2017).

4.2. Business-related barriers

Many scholars draw attention that it is not "the black box issue" which makes the production of knowledge about ML-based systems (or AI in general) a difficult task for regulators. Even in cases where the system in question is simple enough to allow for a proper explanation of the decision, it is rather the legitimate business interests or other institutional concerns, which make individual access to relevant information a delicate balancing act (Wischmeyer 2020, p. 79). The disclosures made for the purposes of contestability may reveal information jeopardizing the integrity of the system or may impair the competitive advantages of the system operator/designer. Thus, it goes without saying that those who deploy ADM systems have a strong interest in the deliberate establishment and maintenance of opacity.[19]

4.2.1. System's integrity

Concealment, nondisclosure, and controlled access are the strategies that data controllers may resort to protect the integrity of their systems by preventing users from gaming or circumventing the decision-making process. Individuals who manipulate the inputs of the system (based on their intimate knowledge of the system's behavior) not only gain advantage for themselves but also impair the predictive capacity of the system. Gaming may be seen as the rational behavior of the users where the cost of manipulating the input is lower than the expected benefits or the eliminated risks.

Gaming of the system – also referred to as adversarial learning in ML context – may involve strategies in the form of *avoidance*, *altered conduct*, *altered input*, and *obfuscation* (Bambauer & Zarsky 2018). Depending on the context and the values that a system designer/operator wants to prioritize, each type of gaming or "gameability" may affect the individual and the society differently, and not necessarily negatively. In cases of *altered conduct*, where the individual changes his/her course of action to avoid adverse effects, the end-result may simply amount to lawful or desired behavior – accomplishing the very objective of the ADM system. This may be seen as a form "nudging," which alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives (Thaler & Sunstein 2008, p. 6).

The gaming behavior may also take the form of *altered input*, which aims to improve some proxy features without actually improving the underlying attributes that the system aims to reinforce. For instance, while a loan applicant may choose to pay her bills on time to increase her credit score, she can also invest in efforts to discover the proxy features and heuristics that she could manipulate to present herself as if she was creditworthy.

(Kleinberg & Raghavan 2018).[20] This is especially the case for ADM systems, which target unobservable or hard-to-measure characteristics and therefore need to use *proxy features* which are assumed to be representative of the actual attributes.

As seen, the policy implications of gaming and its countermoves are a "mixed bag" in that ADM systems may incentivize both productive and unproductive forms of effort. As such, the diversity of the concept resists any overarching theory about how the gaming behavior must be weighed against competing values. Hence, there is no uniform or straightforward justification for secrecy practices aiming to prevent gaming (Bambauer & Zarsky 2018, pp. 22, 33).

Deploying more complex models, making frequent changes in the parameters of the system, and using differently sourced proxies are the strategies employed to make gaming more difficult or less rewarding. Predictably, the systems, which use more immutable characteristics or observe rather nonvolitional behavior are more resistant to gaming. From the legal perspective, data controllers' integrity claims may generally be based on the right to conduct business since the system's accuracy and efficiency diminish due to the "false" data fed by the gaming behavior. In many cases, gaming behavior amounts to a tortious act or a breach of contract. Where appropriate, gaming may also be opposed and counteracted on the basis of public health, privacy, or security risks.

4.2.2. Economic rivalry

Integrity claims are usually conflated with competition-related arguments. Industry players may be reluctant to disclose the coding of the system or the training data, or they may simply refuse to provide an explanation of the ML model as this may weaken their competitive advantage. Intellectual property (IP) rights, and in particular trade secrets,[21] is the main legal framework that businesses rely on to prevent competitors from gaining access to commercially valuable information.

Many informational elements in a ML process such as individual data, databases, algorithms, profiles, data features, or ML models fully or partially fall within the ambit of IP protection. However, speaking of disclosure of and access to ML systems (*1st layer* transparency), only trade secrets may fully be relied upon for the purpose of secrecy. Other types of IP protection do not in principle provide secrecy on the content but focus on the reproduction, dissemination, adaptation, or other specific uses of the protected subject-matter. Hence, copyright (including software protection[22]), *sui generis* database right or patent rights could be relevant in case of *2nd layer transparency* measures (Part V). The procedural, technical, or administrative mechanisms deployed to scrutinize ADM systems or to contest specific decisions may require the copying, reverse engineering, or otherwise modification of the ML elements.[23]

Not much substantial work exists to offer guidance about how different transparency or contestability needs could be reconciled with data controllers' and other industry players' legitimate interests. The issue is also poorly addressed in the GDPR. In connection with the technologies enabling data subjects' remote access to their personal data, Recital 63 of the Regulation reads: "[t]hat right should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software." Although special reference to software may suggest that the statement is confined to remote access systems, an interpretation of the Recital to include Article 22 would contradict neither with the spirit of the GDPR nor with the general principles of law. Having said that, there is also no compelling reason to read too much into this single reference. As with other fundamental rights, the right to property should also be respected in the application of the EU *acquis communautaire* notwithstanding whether there exists a specific reference in the GDPR. Moreover, the wording "adversely affect" in Recital 63 may be seen as no more than a mere reminder because the Recital also assures that "the result of those considerations should not be a refusal to provide all information to the data subject." The WP29 *Guidelines on transparency*[24] further make it clear that businesses cannot rely on trade secret protection as an excuse to categorically deny access or refuse to provide information and recommends a case-by-case approach when dealing with conflicting values and interests.

Irrespective of the affordances of the IP rights, in practice, industry players mostly rely on their physical \control over the systems to keep the ADM systems and the data in the dark. This physical control may also be complemented with contractual terms to prohibit any testing or reverse-engineering of the decision-making process.

5. Beyond impediments: The *2nd layer transparency*

Where the technical limits and/or the business-related concerns prevent human-intelligible contestation, compliance with Article 22 requires the deployment of further tools and methodologies to render ADM systems contestable before a human arbiter and/or by use of software. This Part defines these tools and methodologies as the *2nd layer transparency* measures and explores their compatibility with the GDPR.

5.1. DPbD and implementing transparency under the GDPR

With the explicit reference in the GDPR, many of the *2nd layer transparency* measures (tools and methodologies) fall under the banner of *Data Protection by Design and by Default* (DPbD) as provided in Article 25. DPbD is a generic concept based on the idea that privacy intrusive or other harmful features of a product or service must be limited to what is necessary for the simple use of it. Under Article 25, data controllers are obliged to implement measures *in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects*. Combined with Article 28(1), this requires data controllers and processors to "hard wire" data protection norms into their systems' architecture and *modus operandi* (Yeung & Bygrave 2020). As a "design-based" regulatory technique, the provision targets all personal data processing, irrespective of the technology used to construct and operate the systems.

DPbD entails a series of regulatory, technical, and organizational measures that should be actively followed through the entire life-cycle of data-driven practices. As Article 25 makes a general reference to the "requirements" under the GDPR, it is clear that the DPbD is not limited to implementation of certain data-protection principles but may be extended to a notion of "*Contestability by Design*" (CbD) as a sub-species (Almada 2019). In a similar vein, Hildebrandt and Koops (2010) suggest the concept of "smart transparency," which refers to the designing of the socio-technical infrastructures carrying out automated decisions in such a way that individuals can anticipate and respond to how they are observed or profiled (p. 450).

Inclusion of DPbD as a legal obligation in the GDPR clarifies that the liability as to the design choices and the operational decisions of the ADM system lies with the data controller as the addressee of the norm and thus, may not be shifted to the contractors or third-parties. A further implication of the DPbD is that the design choices should address the rights and obligations outlined in the GDPR (e.g. right to contest), rather than referring to less definable and general notions such as privacy or accountability (Hildebrandt & Tielemans 2013, p. 517). Taking a broad view of "design," Article 25 speaks of "appropriate technical and organizational measures and procedures," which may extend as far as the integration of the relevant methodologies into the business models of data controllers.

5.2. Design choices to reduce complexity for more interpretable results

A variety of design choices might be introduced to enhance or facilitate the contestability of automated decisions as a form *of design-based regulation* aiming to prevent or inhibit certain conduct or social outcomes (Yeung 2015). By interfering with the design and construction stages of the system – rather than addressing its usage or consequences – the idea is to orchestrate the learning process so that the resulting ML model is more amenable to interpretation. *First*, as a rule, systems may be allowed to operate only on a limited set of possible features. By doing so, the total number of relationships handled by the algorithm may be reduced to a human-intelligible level. *Second*, the chosen learning method may allow for models that can be more easily parsed (e.g. decision tree algorithms) in comparison to, for instance, deep learning or neural network type of algorithms. A *third* method could be setting general parameters for the learning process to bring a threshold to complexity so that the resulting model would not defy human comprehension.

In general, *regularization* methods allow for model complexity to be taken into account during the learning process by assigning a cost for excess complexity (Selbst & Barocas 2018, p. 1112). In addition, linear models – with sufficiently small set of features – are regarded to be more concise for humans to grasp the relevant statistical relationships and to simulate different scenarios. Also, systems with *monotonicity* offer simpler models because in monotonic relationships, an increase in an input variable can only result in either an increase or a decrease in the output.

It is usually believed that there is a trade-off between the interpretability and the accuracy of the model. Models considering larger number of variables and more diverse relationships between these variables are assumed to be more accurate. Although this belief is largely relied upon by the system designers when modeling various kinds of problems, it is in fact questionable whether such comparison rests on a rigorous definition of interpretability or conforms to the findings from empirical studies. Hildebrandt notes, the developers may be inclined toward the "low hanging fruit," meaning that they go after the data, which is easily available but not necessarily the most relevant or complete. Yet, more data do not always result in a better target function to define the problem in hand. A detailed model also carries the risk of overfitting and thus weakening its capacity to generalize new data (Hildebrandt 2018, pp. 102, 104). For instance, a model may assign significance to too many features and thus may learn patterns that are peculiar to the training data or not intuitively representative of the phenomena under analysis. Therefore, removing unnecessary features and accordingly reducing the complexity of the model to improve its interpretability does not necessarily decrease the performance of the system (Hand 2006).

Nevertheless, simpler models may not always be expressive enough for proxying sophisticated human behavior. Since ML is best suited to detect subtle patterns and intricate dependencies, it is possible that a complex phenomenon may require a complex model to better account for "reality" (Selbst & Barocas 2018, p. 1129). Article 25 of the GDPR acknowledges this need for "requisite complexity" by setting technical and economic feasibility as the two criteria informing the scope of the obligation of DPbD.[25] This confirms, as a principle, both technical difficulties and economic downsides may set a limit in terms of adopting plainly intelligible decision-making models.

5.3. Procedural, administrative, and institutional measures

As simpler models may not be feasible due to "requisite complexity," there is need for further *institutional*, *administrative*, and *procedural* measures, which may facilitate the monitoring, review, and the contestation of automated decisions. These measures do not aim to impose limits on the ML process (e.g. capping the number of data features) but rather offer ways to improve the accountability and/or interpretability of the system.

Part of these measures fall under the concept of *procedural regularity* which, as a design principle, ensures that ML systems are actually doing what they are declared to be doing by their designers and operators (Kroll *et al.* 2017). In addition to procedural regularity, there are other design features and technical "add-ons" that may be implemented into the system during the development stage. These solutions aim to render ADM systems and their output intelligible to human reason or auditable through algorithmic means (algorithmic scrutiny). For instance, ADM systems may be designed to register processes leading to their actions, identify possible sources of uncertainty, and disclose any assumptions relied upon. In this regard, even a simple internal log kept by the system could enhance transparency for the purpose of contestation. Such records may be arranged to indicate the state of the model at the time of the decision or to provide information about the decisional input together with the rules actually employed for a specific outcome.

Apart from the procedural requirements, various administrative measures may be put to use to improve the accountability and interpretability of the ADM systems. In this respect, the *2nd layer* transparency/contestability measures also accommodate the idea of an institutional setup to carry out the necessary inspection and supervision tasks on behalf of data subjects. Where legitimate secrecy claims of the system designers/operators prevail, institutional oversight may be a way both for ex-ante inspection of the systems and for ex-post challenging of specific decisions (Sandvig *et al.* 2014). Such institutional review allows for "selective transparency," assuring that critical information is not disclosed to the public but kept limited to the legally designated entities representing data subjects (Desai & Kroll 2017).

A number of GDPR provisions either explicitly provide for the above *institutional*, *administrative*, and *procedural* measures or leave space for them. The Recitals of the Regulation – together with several *guidelines* and *recommendations* published by the EU bodies – also articulate requirements and elaborate on the rights and obligations to this effect. In line with this, WP29 *Guidelines on automated decision-making* refers to *code of conduct* (Article 40, Recitals 77 and 98), *certification* (Article. 42), *agreed standards*, and *ethical review boards* as

formal mechanisms for scrutinizing ADM. Yeung and Bygrave (2020) regard these *self-*, *meta-*, and *coregulatory* instruments and techniques as a *cooperative problem-solving approach* between the regulator and the regulatee.

Though not mandatory under the GDPR, certification – which verifies that a product, process, or service adheres to a given set of standards and/or criteria – may require the disclosure of extensive technical information. This may include the source code, the hardware/software environments in which the systems has been developed, and the performance of the system in the testing environments (Hoffmann-Riem 2020, p. 13). As an integral part of the certification process, the International Organization for Standardization (ISO) has so far published several standards on *big data*, and further developing more on AI. *Institute of Electrical and Electronics Engineers Standards Association* (IEEE-SA) is also working on ML standards such as P7001, which focuses on transparent operation of autonomous systems. According to the 2019 report (*Ethically Aligned Design*) of the IEEE, the aim is to describe measurable and testable levels of transparency so that autonomous systems can be objectively assessed to determine their levels of compliance. Regarding these efforts, Matus and Veale (2020) note that, rather than prescribing concrete formulations, standard-setting initiatives have so far been limited to terminology issues and analytical frameworks laying out "meta-standards."

Data Protection Impact Assessment (DPIA), as provided in Article 35 of the GDPR, could also serve as an important transparency mechanism which could aid the scrutiny of ML-based systems in various ways and dimensions. It is argued that a version of an Algorithmic Impact Assessment (AIA) might be derived from the DPIA to obtain an external review of the system. WP29 *Guidelines on automated decision-making* confirms this by mandating DPIA for any ADM subject to Article 22. In a similar vein, Kaminski and Malgieri (2020) approach DPIA as a "collaborative governance mechanism," which may help determine the optimum extent of transparency and the appropriate mode of implementation of the right to contest in a specific ADM context (p. 72).

Despite these regulatory and institutional affordances provided by the GDPR together with various soft-law documents, it is arguable whether the current data protection authorities both at the member state and the Union level can handle such a wide range of regulatory, monitoring, and auditing tasks. As this will require a significant expansion of their powers and personnel, there are also views in favor of specialized institutions to monitor AI-based applications and develop performance standards (Hoffmann-Riem 2020, pp. 14–15).

5.4. Algorithmic scrutiny

There exist a variety of algorithmic scrutiny tools that enable both ex-ante and ex-post testing and verification of ADM processes. The deployment of these tools for the practical purpose of scrutinizing automated decisions present a spectrum ranging from modules integrated into the systems to stand-alone external audit tools for "black-box testing" (Pedreschi *et al.* 2018). Any combination of these may also, in varying degrees, involve humans in the decision loop to adjust the specifications and interpret the results.

Technologies for algorithmic scrutiny may be used both to approximate the ML model in general and also to discover the features that are most relevant for an individual decision. The former, a.k.a. "global interpretability," aims to understand the underlying logic and the mode of reasoning of the system in its entirety (Selbst & Barocas 2018, p. 1113). Global interpretability can be regarded to generate a *model of the model* to simulate possible outcomes. The idea is to reconstruct the model on the basis of *interpretable rules* that describe the input–output relationships. "Local interpretability" on the other hand, is an ex-post method looking for the reasons for a specific decision. It is the "review of a software-driven action after the fact of the action" (Desai & Kroll 2017, p. 39). The local interpretability tools generally focus on importance-measuring methods aiming at explaining the most important variables for a specific result. It is a user-centric approach, where the importance of any feature to a particular decision is detected by iteratively varying the value of that feature, while holding the others constant. The idea is to develop an interpretable model taking on the predictions of a supposedly uninterpretable (black-box) model. This enables to determine the relative contribution of different features and identify the values that need to be altered to bring about a certain outcome (Ribeiro *et al.* 2016). Although local interpretability seems to work well to explain a specific decision, solely employing these limited techniques without a model-centric inspection or verification may be misleading. This is mainly due to the fact that an explanation that accounts

for a certain decision does not apply in the same way to other decisions (Doshi-Velez & Kortz 2017). That is, the reasons for a specific decision do not illustrate a general rule with regard to the system's behavior and thus, may be insufficient for the purposes of contesting another decision. Especially in terms of understanding the context of the decision, a proper scrutiny of automated decisions requires both the use of system-centric and user-centric approaches simultaneously.

Although ex-post techniques may be used as stand-alone scrutiny tools, their success depends on the extent to which they are reinforced by the necessary administrative, technical, and organizational measures in an overarching DPbD framework. As Article 25(1) clearly states that DPbD should be pursued through adequate technical and organizational means (*both at the time of the determination of the means for processing and at the time of the processing itself)*, the provision may be interpreted to include the construction of software-based tools enabling contestation.

In addition, Recital 71 of the GDPR on automated decisions is also found to be supportive of software tools for the purpose of review and oversight. The Recital states that the data controllers are under the duty to implement technical and organizational measures against inaccuracies in data and to minimize the error against the risks involved for the rights and interests of the data subject. The WP29 *Guidelines on automated decisions* also recommends "algorithmic auditing" as a safeguard under Article 22. However, rather than the scrutiny of individual decisions, these references envisage algorithmic audit as a general model-centric tool to assess data controllers' compliance. The strongest support for the algorithmic contestation of specific decisions may be found in the wording of Article 22 itself. The provision defines the right to obtain human intervention as the least of the measures that the data controller could implement – implying that further solutions such as software-based tools could also be necessary for the purposes of contestation. Lastly, Article 21(5) on the right to object to personal data processing, including profiling, provides that "the data subject may exercise his or her right to object by automated means using technical specifications."

6. Options for implementation

Having seen the possible legal, technical, and organizational measures to overcome the opacities and the legal barriers that may stand in the way of contestability of the automated decisions, it becomes clear that there is no one-size-fits-all solution (Kaminski 2019a). The problem lays with the framing of the optimum extent of transparency and the appropriate mode of implementation without prejudice to the integrity of the systems or to the legitimate interests of the stakeholders involved.

Following the analysis in the previous Parts, what remains yet to be answered is the question of *which uses of ADM should be subject to right to contest, and through which (combination) of the above tools and measures?* As this paper is primarily focused on a conceptual analysis of the right to contest and the relevant affordances provided by the GDPR, further enforcement-related particularities are beyond the scope. Therefore, this Part is limited to an outline of the possible regulatory options regarding the implementation of the right to contest.

6.0.1. Not permissible

Where satisfactory measures for the exercise of the right to contest at human-intelligible level are not possible or feasible, taking into account the potential risks to the rights of the data subjects or to the societal interests, solely automated decisions may be banned. In case the benefits from the ADM do not justify the risks that it creates, system developers and operators should look for hybrid (human-machine symbiotic) approaches.

In their implementation of the GDPR, many member states have introduced prohibitions or restrictions for certain types of decisions. Among them, French Law adopts a regime based on traditional state functions and accordingly prohibits fully or semiautomated *judicial decisions* aiming to evaluate aspects of personality, while permitting automated *administrative decisions* subject to strict conditions (Malgieri 2019, p. 13–14). The exclusion of certain types of automated decisions also means that data protection principles shall be respected at all stages of ADM, including the initial decision on whether or not to carry out the processing.

6.0.2. Permissible – Subject to ex-ante design and procedural requirements and/or ex-post algorithmic scrutiny

Despite the methodological distinction we have made in Part IV, transparency issues are never of purely technical, legal, or institutional nature. The complexity-related problems – that are deeply intertwined with both

deliberate and unintentional design features – often result in unintelligible ML models. Therefore, determining the right combination of the *2nd layer* transparency measures requires a context-specific case-by-case approach, taking into account the technical limits, possible gaming strategies, and the competition-related issues. While in some cases only an ex-post analysis (black-box testing) may suffice, the *institutional*, *administrative*, or *procedural* measures are more effective when accompanied with ex-post tools and methodologies.

6.0.3. Permissible – Only subject to ex-post black-box testing

This regulatory option deals with the situations, where it is not possible or permissible to impose constraints on the model or apply other ex-ante measures at the design stage. In such cases, data controllers may still be allowed to carry on with ADM subject to ex-post algorithmic scrutiny measures. Since black-box testing alone may not reveal sufficient insights about the decision-making process, this type of scrutiny may remain limited to testing of the outcome against some minimum (e.g. fairness, antidiscrimination, or due process) requirements, without involving a full-fledged contestation. Considering the risks involved, there may be need for further procedural safeguards such as the immediate suspension of the automated decision upon challenge or the reversed burden of proof in the contestation proceedings.

6.0.4. Permissible without restrictions – Only subject to 1st layer access and notification requirements

This is where the individual access rights and notification duties under the GDPR suffice to provide functional and systemic information about the input, the decision rules, and the underlying causal relations necessary for contestation.

7. Conclusion

The implementation of the above options (modalities) as a practically meaningful transparency scheme requires deeper interdisciplinary research on two parallel but interacting tracks. *First*, there is need for an elaboration of the technical limits and other impediments to human-intelligible models. That is, where there are genuine technical and legal barriers and where complexity and/or legal claims are used as a pretext for unsubstantiated or unlawful secrecy practices. This line of inquiry, briefly touched upon in Part IV, will involve the consideration of heterogeneous composition of legal matters including fundamental rights, legislative initiatives, case-law, data protocols, regulatory tools, contracts, and so on. Without understanding the true nature and the cause of intransparencies in a given system, it will not be possible to calibrate the measures to be implemented. It is hoped that the analysis provided in this paper could serve as a "launching pad" for further legal research to draw an entire picture of legal impediments in relation to the implementation of the right to contest. As we identify what counterbalancing rights and interest are at stake on the side of the data controllers and other industry actors, the *second* line of research should then inquire how to define and treat the risks to the rights of the data subjects. Risk assessment for the practical application of the above regulatory options should initially identify the conditions where a less interpretable model – for the sake of efficiency gains or other alleged benefits – could be justified. This would require an overall consideration of the type and source of the data, the reliability of the ML model, the specific conditions of the data subject, and most importantly, the materiality of the output to the individuals and third parties concerned.

Despite the flux of academic papers in the recent years about the *transparency*, *explainability*, *interpretability*, *legibility*, and the discriminative and unfair effects of automated decisions, we are still far from establishing any practical way of exercising rights under Article 22. To bridge the gaps between the GDPR and its practical application, the affordances laid out in this paper need to be operationalized as a comprehensive compliance scheme. Failure to do this uniformly at the EU level may result in a fragmented implementation, rendering legal safeguards to a great extent ineffective (Massé & Lemoine 2019). That is, despite the proactive approach of some member states such as Hungary and France, for many member states, Article 22 may remain an ancillary provision as it has been during the time of the DPD (Malgieri 2019). Considering the current political incoherence among the EU member states, it would not be unrealistic to expect further disarray with regard to the implementation of Article 22 if Member States are left to legislate on their own discretion.[26]

Whether or not the GDPR provisions on automated decisions will turn out to be a toothless mechanism depends on whether the EDPB and other EU authorities take prompt action. In this respect, there is need for an

agenda for the development of a dynamic and scalable compliance regime with concrete practical targets. The progress on this front will determine whether Bygrave (2001) is still right in his conclusion about Article 15 of the DPD (GDPR, Art. 22), which he had phrased 20 years ago as: "all dressed up but nowhere to go."

Acknowledgments

The author received funding from Tilburg Institute for Law, Technology, and Society (LTMS-TILT), Tilburg University, The Netherlands, and the paper was finalized during a research stay at the Institute of Computing and Information Sciences (iCIS), the Science Faculty of Radboud University, The Netherlands, funded by the Privacy and Identity Lab.

Endnotes

[1] Similar transparency obligations (requirements) are emerging in various legal domains and regulatory frameworks, for example, *Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services*. The Regulation foresees disclosure obligations for the providers of online intermediation services which use algorithms to rank goods and services. See Wischmeyer 2020, p. 77.

[2] For a literature review shedding light on the diversity of the concept of *algorithmic regulation*, see Eyert *et al*. 2021.

[3] Approaching transparency in ADM as a procedural problem also aligns with the process-based nature of the EU data protection regime, which has long focused on the regulation of digital interactions while maintaining regulatory flexibility in the face of complex and ever-changing technology (Yeung & Bygrave 2020).

[4] For different conceptions of transparency under the GDPR, see Felzmann *et al*. 2019. Authors propose to understand transparency relationally, where information provision is conceptualized as communication between technology providers and users, and where assessments of trustworthiness based on contextual factors mediate the value of transparency communications.

[5] For this "layered" approach, see Bayamlıoğlu "Transparency of Automated Decisions in the GDPR: An Attempt for Systemization" (PLSC 2018 Discussion Paper, Brussels). Kaminski also speaks of "tiers" of transparency under the GDPR, albeit in a different conception (Kaminski 2019a). Diakopoulos and Koliska (2017) use the "layers" concept to define various aspects of disclosable information in relation to a "transparency model". Also *see* note 15.

[6] The origin of these provisions is attributed to French data protection legislation enacted in 1978, which prohibited behavioral assessment through automated processing in legal matters (Bygrave 2001, p. 21; Korff 2010, p. 83).

[7] Hildebrandt notes, albeit critically, "[t]he fact that usually some form of routine human intervention is involved means that art. 15 is not applicable, even if such routine decisions may have the same result as entirely automated decision making" (Hildebrandt 2008, p. 28, fn. 22).

[8] EC Commission's amended proposal of DPD 1992. COM(92) 422 final – SYN 287, 15.10.1992, 26.

[9] Here, the question arises whether *consent* as used in Article 22 has the same requirements as *consent* defined in Articles 3 and further regulated in Art. 7 of the GDPR.

[10] For instance, WP29 *Guidelines on automated decisions* is also silent on this matter. The Guidelines refer to the safeguards cumulatively as "a further layer of protection for data subjects" (p. 15).

[11] This procedural approach to Article 22, treating DP law as meta-regulation (a body of rules, regulating a regulatory technology), partly addresses the arguments that DP regime slowly becomes the "law of everything," penetrating every sector where digital technologies are involved, see Purtova 2018.

[12] "Normativity" is hereby used in the sense of not being explicable in purely factual terms. Such attribution does not make sense for ordinary physical objects, such as rocks, geological systems, or oil molecules—leaving aside living organisms.

[13] The transparency implications (contestability requirements) in this Section partially parallels with Diakopoulos and Koliska's (2017) "transparency model," which enumerates *information factors* that might be disclosed about algorithms. Their model provides a set of pragmatic dimensions of information (across layers such as data, model, inference, and interface) that are essential for algorithmic transparency efforts.

[14] See WP29 2018 *Guidelines on automated decisions*, raising the questions about the "categories of data used in the profiling or decision-making process" and "why these categories are considered pertinent" in relation to the implementation of Article 22. In addition, Article 14 of the GDPR clearly counts "categories of personal data" among the information to be communicated to the data subject, though without guidance about what these categories might be other than sensitive/

nonsensitive data (also see Article 28 and 30). On this regulatory gap regarding the categories of personal data, see Wachter *et al.* (2018).

[15] *Jehovan todistajat*, C-25/17, EU:C:2018:551.

[16] The case discussed the joint controllership between the website owner, *Fashion ID*, and *Facebook* Ireland, where the former has embedded in its website the Facebook "Like" button. Case C-40/17, *Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW eV*. ECLI:EU:C:2019:629.

[17] *WP29 Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*' (last revised and adopted on 6 February 2018). As of 25 May 2018, the Article 29 Working Party (WP29) ceased to exist and is replaced by the EDPB.

[18] The "technical limits," as explained here, partly corresponds with the concept of "epistemic opacity" that Evert and others define as "inherent methodological intransparency of ML" (Eyert *et al.* 2021). Also see Wischmeyer referring to "epistemic constraints" (Wischmeyer 2020, p. 80).

[19] Evert and others conceptualize these informational asymmetries as "sociomaterial opacity" arising due to the concentration of massive data sets in the hands of a few private companies as well as the inaccessibility of closed-source algorithms (Eyert *et al.* 2021).

[20] On the question of when algorithms need to be kept secret due to the risk of gaming and when disclosure is permissible, see Cofone & Strandburg 2019.

[21] *European Parliament and the Council, Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure* (OJ L 157/1, 15.6.2016), 8 June 2016.

[22] *Directive 2009/24/EC on the legal protection of computer programs* (Computer Programs Directive), (OJ L 111/16, 5.5.2009).

[23] On this matter, see Benjamin L. W. Sobel, "Artificial Intelligence's Fair Use Crisis" 41 COLUM. J.L. & ARTS 45 (2017); Banterle (2018); Mattioli (2014).

[24] *WP29 'Guidelines on transparency under Regulation 2016/679 (WP260)'* (EC, 24 January 2018)

[25] It should also be noted that "whatever seemed technically and/or economically infeasible during the design of the data processing system, will again be considered once the processing is in operation" (Hildebrandt & Tielemans 2013, p. 517).

[26] Malgieri provides a detailed account of the current implementation efforts of the EU member states. His analysis reveals that member states have so far taken various approaches mostly without a clear or concrete methodology. Apart from some good examples, member state laws generally rephrase the wording of the official documents or refer to "explanation" in an abstract and descriptive way (Malgieri 2019).

References

Almada M (2019) Human Intervention in Automated Decision-Making: Toward the Construction of Contestable Systems. ICAIL '19: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, June 2019 pp. 2–11. https://doi.org/10.1145/3322640.3326699

Baer T (2019) *Understand, Manage, and Prevent Algorithmic Bias: A Guide for Business Users and Data Scientists*. Apress Imprint, Berkeley, CA.

Bambauer J, Zarsky T (2018) The Algorithm Game. *Notre Dame Law Review* 94(1), 1–48.

Banterle F (2018) The Interface between Data Protection and IP Law. In: Bakhoum M *et al.* (eds) *Personal Data in Competition, Consumer Protection and Intellectual Property Law*, pp. 412–440.Springer-Verlag GmbH, Germany.

Bateson G (1987) *Steps to an Ecology of Mind*, (1972 1st edition). Jason Aronson Inc, Northvale, NJ, and London.

Bayamlıoğlu E (2018) Contesting Automated Decisions. *European Data Protection Law Review* 4(2018), 433–446.

Bayamlıoğlu E, Leenes R (2018) The Rule of Law Implications of Data-Driven Decision-Making: A Techno-Regulatory Perspective. *Law, Innovation and Technology* 10, 295–313.

Binns R (2017) Algorithmic Accountability and Public Reason. *Philosophy & Technology* 31, 543–556.

Burrell J (2016) How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* 3 (1), 1–12.

Bygrave L (2001) Automated Profiling: Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling. *Computer Law & Security Review: The International Journal of Technology Law and Practice* 17, 17–24.

Castañeda HN (1970) On the Semantics of the Ought-to-Do. *Synthese* 21, 449–468.

Cath C (2018) Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges. *Philosophical Transactions of the Royal Society A* 376, 20180080. https://doi.org/10.1098/rsta.2018.0080.

Cofone I, Strandburg K (2019) Strategic Games and Algorithmic Secrecy, 64 McGill L.J.

Commission, E., Directorate General Justice, F. & Security (2010). Comparative Study on Different Approaches to New Privacy Challenges in Particular in the Light of Technological Developments. Working Paper No. 2. In: D. Korff (ed.), Data

Protection Laws in the EU: The Difficulties in Meeting the Challenges Posed by Global Social and Technical Developments, London Metropolitan University, London.
De Ridder J (2006) The Inherent Normativity of Technological Explanations. *Techné: Research in Philosophy and Technology* 10(1), 79–94.
Desai D, Kroll J (2017) Trust but Verify: A Guide to Algorithms and the Law. *Harvard Journal of Law & Technology* 31, 2–64.
Diakopoulos N, Koliska M (2017) Algorithmic Transparency in the News Media. *Digital Journalism* 5(7), 809–828. https://doi.org/10.1080/21670811.2016.1208053.
Doshi-Velez F, Kortz M (2017) Accountability of AI under the Law: The Role of Explanation. Harvard University Berkman Klein Center Working Group on Explanation & the Law, Working Paper No. 18-07. http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372584
Economides C (2010) Content of the Obligation: Obligations of Means and Obligations of Result. In: Crawford J, Pellet A, Olleson S, Parlett K (eds) *The Law of International Responsibility*. Oxford University Press, New York.
Edwards L, Veale M (2017) Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking for. *Duke Law & Technology Review* 16, 18–84.
Eyert F, Irgmaier F, Ulbricht L (2021) Extending the Framework of Algorithmic Regulation. The Uber Case. *Regulation & Governance* (in this issue).
Felten E (2017) What Does It Mean to Ask for an "Explainable" Algorithm? *Freedom to Tinker*. Available from URL: https://freedom-to-tinker.com/2017/05/31/what-does-it-mean-to-ask-for-an-explainable-algorithm/
Felzmann H, Villaronga EF, Lutz C, Tamò-Larrieux A (2019) Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns. *Big Data & Society* 6, 1–14. https://doi.org/10.1177/2053951719860542.
Gollnick C (2018) Induction Is Not Robust to Search. In: Bayamlıoğlu E, Baraluic I, Janssens L, Hildebrandt M (eds) *Being Profiled Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, pp. 106–111.Amsterdam University Press, Amsterdam.
Guinchard A (2017) Contextual Integrity and EU Data Protection Law: Towards a More Informed and Transparent Analysis. *SSRN Electronic Journal*, 2017. https://dx.doi.org/10.2139/ssrn.2946772.
Hand D (2006) Classifier Technology and the Illusion of Progress. *Statistical Science* 21, 1–14.
Hildebrandt M (2008) Defining Profiling: A New Type of Knowledge. In: Hildebrandt M, Gutwirth S (eds) *Profiling the European Citizen Cross-Disciplinary Perspectives*, pp. 17–45.Springer, Dordrecht.
Hildebrandt M (2018) Preregistration of Machine Learning Research Design. Against P-Hacking. In: Bayamlıoğlu E, Baraluic I, Janssens L, Hildebrandt M (eds) *Being Profiled Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, pp. 102–105.Amsterdam University Press, Amsterdam.
Hildebrandt M (2019) Privacy as Protection of the Incomputable Self: From Agonistic to Agnostic Machine Learning. *Theoretical Inquiries of Law* 19(1), 83–121.
Hildebrandt M, Koops J (2010) The Challenges of Ambient Law and Legal Protection in the Profiling Era. *Modern Law Review* 73(3), 428–460.
Hildebrandt M, Tielemans L (2013) Data Protection by Design and Technology Neutral Law. *Computer Law & Security Review* 29(5), 509–521.
Hoepman J (2018) Transparency as Translation in Data Protection. In: Bayamlıoğlu E, Baraluic I, Janssens L, Hildebrandt M (eds) *Being Profiled Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, pp. 102–105.Amsterdam University Press, Amsterdam.
Hoffmann-Riem W (2020) Artificial Intelligence as a Challenge for Law and Regulation. In: Wischmeyer T, Rademacher T (eds) *Regulating Artificial Intelligence*, pp. 1–32.Springer, Cham.
Jasanoff S, Simmet H (2017) No Funeral Bells: Public Reason in a "Post-Truth" Age. *Social Studies of Science* 47(5), 751–770.
Kaminski M (2019) The Right to Explanation, Explained. *Berkeley Technology Law Journal* 34(1), 189. https://scholar.law.colorado.edu/articles/1227.
Kaminski M (2019a) Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability. *Southern California Law Review* 92(6), 1–77.
Kaminski M, Malgieri G (2020) Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 68–79. https://doi.org/10.1145/3351095.3372875.
Kleinberg J, Raghavan M (2018) How Do Classifiers Induce Agents To Invest Effort Strategically? arXiv: 1807.05307v5
Krist V (2006) How Norms in Technology Ought to Be Interpreted. *Techné: Research in Philosophy and Technology* 10(1), 95–108. https://doi.org/10.5840/techne200610144.
Kroll JA (2018) The Fallacy of Inscrutability. *Philosophical Transactions of the Royal Society A* 376(2133). http://doi.org/10.1098/rsta.2018.0084.
Kroll JA, Huey J, Baroas S *et al.* (2017) Accountable Algorithms. *University of Pennsylvania Law Review* 165(3), 633–705.
Lipton Z (2016) The Mythos of Model Interpretability, *ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*, New York. [Last accessed 25 Jun 2019.] Available from URL: https://arxiv.org/pdf/1606.03490.pdf
Macmillan R (2018) *Big Data, Machine Learning, Consumer Protection and Privacy*. Geneva, Switzerland: International Telecommunication Union (ITU) Security, Infrastructure and Trust Working Group.
Malgieri G (2019) Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations. *Computer Law & Security Review* 35(5), 105327.
Malgieri G, Comandé G (2017) Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation. *International Data Privacy Law* 7(4), 243–265.

Mantelero A (2019) Artificial Intelligence and Data Protection: Challenges and Possible Remedies. The Council of Europe's Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6

Massé E, Lemoine L (2019) One Year Under the GDPR, *Access Now Publication*. Available from URL: https://www.accessnow.org/cms/assets/uploads/2019/06/One-Year-Under-GDPR.pdf

Matus KJM, Veale M (2020) Certification Systems for Machine Learning: Lessons from Sustainability. *Regulation & Governance* (in this issue).

Mendoza I, Bygrave L (2017) The Right Not to Be Subject to Automated Decisions Based on Profiling. In: Synodinou T, Jougleux P, Markou C, Prastitou T (eds) *EU Internet Law*, pp. 78–98.Springer, Cham.

Pedreschi D, Giannotti F, Guidotti R, Monreale A, Pappalardo L, Ruggieri S, Turini F (2018) Open the Black Box Data-Driven Explanation of Black Box Decision Systems. arXiv: 806.09936v1.

Purtova N (2018) The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law. *Law, Innovation and Technology* 10, 74–75.

Rader E, Cotter K, Cho J (2018) Explanations as Mechanisms for Supporting Algorithmic Transparency. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pp. 1–13. Montréal, QC, Canada: ACM Press. https://doi.org/10.1145/3173574.3173677.

Ribeiro MT, Singh S, Guestrin C (2016) "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining–KDD '16*, pp. 1135–1144. San Francisco, CA: ACM.

Sandvig C, Hamilton K, Karahalios K, Langbort C. (2014) Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. Presented at *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*. 22 May, Seattle, WA.

Selbst A, Barocas S (2017) Regulating Inscrutable Systems Available from URL: http://www.werobot2017.com/wp-content/uploads/2017/03/Selbst-and-Barocas-Regulating-Inscrutable-Systems-1.pdf

Selbst A, Barocas S (2018) The Intuitive Appeal of Explainable Machines. *Fordham Law Review* 87, 1085–1139.

Thaler RH, Sunstein CR (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Wachter S, Mittelstadt B, Floridi L (2017) Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law* 7(2), 76–99.

Wachter S, Mittelstadt B, Russell C (2018) Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology* 31(2), 841–887.

Wischmeyer T (2020) Artificial Intelligence and Transparency: Opening the Black Box. In: Wischmeyer T, Rademacher T (eds) *Regulating Artificial Intelligence*, pp. 75–101.Springer, Cham.

Yeung K (2015) Design for Regulation. In: van den Hoven J, van de Poel I, Vermaas PE (eds) *Handbook of Ethics, Values and Technological Design*, pp. 447–472.Springer, Dordecht.

Yeung K (2018) Algorithmic Regulation: A Critical Interrogation. *Regulation & Governance* 12(4), 505–523.

Yeung K, Bygrave L (2020) A Critical Examination of the Legitimacy of the Modernised European Data Protection Regime Through a "Decentred" Regulatory Lens. *Regulation & Governance*, forthcoming.

Zarsky T (2017) Incompatible: The GDPR in the Age of Big Data. *Seton Hall Law Review* 47, 995–1020.

Zeng Z, Fan X, Miao C, Wu Q, Cyril L (2018) Context- Based and Explainable Decision Making with Argumentation. *AAMAS '18 Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, 10–15 Jul, Stockholm, pp. 1114–1122. Available from URL: http://ifaamas.org/Proceedings/aamas2018/pdfs/p1114.pdf

Chapter 5 (*Fourth paper*)

ML and the relevance of IP rights: An account of transparency requirements for AI

• Emre Bayamlıoğlu, "ML and the relevance of IP rights : An account of transparency requirements for AI", *European Review of Private Law Special Issue,* 2023 (forthcoming)
.

I think the term "intellectual property" should be avoided, not because it's a bad term, but because it mixes things up that shouldn't be mixed up. There are different forms, and they hardly have anything to do with each other. [*]

[*] An Interview With Linus Torvalds, Creator of Linux, by Dylan Love, Slate https://slate.com June 09, 2014

Machine learning and the relevance of IP rights (with an account of transparency requirements for AI)

Abstract

As a sub-branch of Artificial Intelligence, Machine Learning (ML) is an inductive method of problem solving which can accomplish tasks that once required human participation and discretion. As governments and other institutions increasingly deploy ML-based systems to predict, rate and act upon individuals' behaviour or personal traits, there is growing political and legal demand for transparency so that the outcome of these systems could be interpretable, and thus contestable where necessary.

Previous research has revealed that transparency in automated decision-making entails not only openness and disclosure in the conventional sense but further administrative and technical measures such as the algorithmic audit or black-box testing of these systems. The implementation of such broadened scope of transparency inevitably involves the reproduction and/or adaptation of the relevant informational elements and components of the ML-based systems. This gives rise to the questions: i) to what extent reliance on Intellectual Property (IP) rights could excuse automated decision-makers from the obligation of making transparent and contestable decisions, e.g., under Article 22 of the GDPR; ii) what are the counter-arguments based on statutory exceptions and limitations restricting IP rights; and iii) what may be the possible solutions either within the IP regime or through regulatory intervention. Overall, the paper aims obtain a macro-view of the potential areas of conflict between the possible transparency measures/tools and the relevant IP regimes—i.e., copyright, sui generis database right and trade secret protection.

1. Introduction

As companies, governments and other institutions increasingly deploy machine learning(ML)-based systems to predict, rate and act upon individuals' behaviour or personal traits, there is growing political and legal demand for transparency (e.g., GDPR Art. 22) so that the outcome of these systems could be interpretable, and thus contestable where necessary.[1] Article 22 and other regulatory attempts aiming for "opening the black-box"[2] brings in the need to develop capacities to monitor *automated decision-making* (ADM) systems and to challenge their unlawful or illegitimate outcomes.[3]

In this vein, previous research has revealed that contesting automated decisions entails not only openness and disclosure in the conventional sense but further administrative measures and technical tools deployment of which may involve the copying, transforming or adaptation of the parts of the ML systems or the relevant datasets. Since such broadened scope of transparency raises the question of possible infringement of intellectual property (IP*)* eligible elements contained in ML systems, this paper provides a systematic analysis of the potential areas of conflict between the IP regime and the transparency requirements implicit in the right to contest.

Following the introduction, Part II briefly examines what should be made visible or intelligible about the ML systems (to render them contestable) and explains the IP relevance of these contestability (transparency) requirements. As such, Part II lays out the structure of the IP analysis which examines ML systems in a dual ontology as: data and datasets (expressional elements) and algorithmic techniques and ML models (operational/functional elements).[4] Having set this background, the rest of the paper inquires to what extent those who deploy ADM systems could rely upon the IP rights to mitigate their transparency obligations.

Part III focuses on copyright and *sui generis* database protection of data and datasets (databases) in a tripartite taxonomy as: i) the training/test data, ii) the "actual" data analysed for a specific decision, iii)

---

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ 2016 L 119/1 ('GDPR') Art. 22 of the GDPR provides that where such technology is used to make decisions (with legal or similarly significant effects for the individuals) without any human oversight, data subjects shall have the right to obtain human intervention and to contest the decisions—as the least of the safeguards that the data controller is obliged to offer.

[2] ML-based systems are frequently referred to as "black-box" for their obscure and impenetrable internal workings which can only be observed in terms of their inputs and outputs. The term is used as a generic metaphor to indicate the seeming complex, inscrutable and opaque nature of AI systems. *See* Frank Pasquale, *The black box society* (Cambridge: Harvard University Press 2015). Also *see* Taina Bucher, *If...Then: Algorithmic Power and Politics* (Oxford: Oxford University Press 2019), 42-46.

[3] Similar transparency requirements are emerging in various legal domains and regulatory frameworks, e.g. Council Regulation (EC) 2019/1150 of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services [2019] OJ L 186. The Regulation foresees disclosure obligations for the providers of online intermediation services which use algorithms to rank goods and services. For regulatory efforts in Germany, see Thomas Wischmeyer,"Artificial Intelligence and Transparency: Opening the Black Box" in Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence* (Cham, Switzerland: Springer 2020), 75-101, 77.

[4] The term *expressional elements*, as used in this paper, refers to the creative literal dimension of the protected subject -matter of copyright protection as opposed to function. dimension.

data comprising of ML output (predictions, ratings etc). Regarding copyright protection, first, the individual data items (text, audio, video etc) in a database may be eligible to copyright protection under the InfoSoc Directive.[5] Second, databases (as compilations) may enjoy protection due to the creativity in the selection and the arrangement of the items comprising them (Art.3(1) of the EU Database Directive). Other than copyright, *sui generis* database protection under the EU Database Directive is another major IP regime which is relevant for algorithmic transparency. Accordingly, Part III examines under which conditions *extraction from* or *transformation of* databases to scrutinise ML systems could amount to an infringement of the *sui generis* database right. The analysis further extends to whether machine generated data could pass the substantial investment test regarding the obtaining, verification or presentation of the contents of the database as provided by the Directive. This part closes with an assessment of the newly enacted text and data mining (TDM) exception in order to see to what extent the exception could facilitate the transparency demands.[6]

Part IV is reserved for the IP eligibility of utilitarian (functional) components and elements of ADM systems as: *algorithms (algorithmic techniques)*, *ML models* and the implementing *computer code*. The analysis is based on the perspective that the enactment of transparency mechanisms such as the software tools for audit and testing, may necessitate the reverse engineering or the implementation (thus reproduction) of the computer code or the essential parts of the system. In Part IV, these IP implications are discussed in relation to both copyright (granted for the literary elements of computer programs) and patent law (where ML-based systems, or parts of them, qualify as novel inventions).

The final leg of the IP analysis, Part V, inquires how *trade secret* protection could act as a barrier against transparency demands. As confidentiality is a highly appealing mechanism for securing ML-based systems, in many cases, relying on trade secrets is the preferred legal strategy of the system designers and operators. The analysis clarifies that both the *expressional* and the *operational* elements of ML systems could satisfy the secrecy requirement under the EU Trade Secret Directive.

The paper concludes that the transparency measures to render ML-based decision systems interpretable and contestable do not easily fit in the exceptions and limitations provided in the relevant IP regimes and discusses further legal approaches which could provide the optimum extent of transparency with minimum prejudice to the integrity of the systems or to the legitimate interests of the stakeholders involved.

2. Contesting automated decisions: An overview of transparency requirements and the relevance of IP rights

2.1 Transparency for the purposes of contesting automated decisions

As a sub-branch of Artificial Intelligence (AI), ML is a method of problem solving increasingly deployed to accomplish tasks which once required human participation and discretion. In order to do that, a data scientist tests various mathematical functions with the aim to reach the formula which best describes the training data for the given purpose(s). The success of ML is dependent on the need to

[5] Council Directive 2001/29/EC of 22 May 2001 concerning certain aspects of copyright and related rights in the information society [2001] OJ L 167 ('InfoSoc Directive').
[6] Articles 3 and 4 of the Council Directive 2019/790/EC of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Council Directives (EC) 96/9 and 2001/29 [2019] OJ L 130 ('DSM Directive').

exploit a critical mass of training data to discover sufficiently descriptive patterns. As an inductive method progressing from particular cases, ML-based systems accumulate a set of discovered dependencies, correlations or relationships that are referred to as the "ML model". [7] A ML model, relying on past examples, develops rules that enable it to perform a predictive or decisional task (e.g., whether past professional conduct is a good predictor of tax fraud or whether those who buy beer also buy diapers and *vice-versa*).

As computational processes, ML-based decisions consist of three main phases: *input, operation* and *output*. A program reads in data, executes an algorithm, and writes out new data as the outcome.[8] Accordingly, contesting an automated decision on normative grounds comes along with a reconstruction of the outcome as a process where certain input leads to certain results— akin to the decisions in a legal system based on *facts*, *norms* and the ensuing consequences. In sum, effective contestation of automated decisions entails procedures and mechanisms enabling insights into the below elements and aspects of the ML system.[9]

1) The "actual data" used for a specific decision: In order to contest a ML-based decision, one initially needs to know the type and the source of the data used (e.g., biometric, health, social media, financial, search engine query, geo-location, energy grid, smart phone use and so on). In addition, the methods for gathering, generating and aggregating data are also essential because data is always prefigured by an interpretive frame and by the technological means that are available.[10] Algorithms and ML models are only as good as the input they are fed with. If input values are false, fraudulent, skewed, or ignored, the resulting output will not be efficient as a problem solver.

2) What the system learns or predicts: It is further necessary to know what type of inferences (facts/evidence) will be harvested from the aggregated data, and eventually fed into the decision *process* of the system.[11] This may relate to one's sentimental mood, health condition, age, physical features, racial indicators, educational level, creditworthiness, income, family relations, sexual preference, work performance and so on. The information to be provided should give an overall picture of how the world is perceived by the system and what "realities" are constructed about the phenomena under analysis.

3) The outcome (decision) and the purpose of the decision-making process: Contestation will further require the knowledge of how the ML outcome serves for the purpose of decision-making in a given context. Combining predictive results with a chosen action, ML systems subtly lay out norms and rules (regulatory frameworks) that affect individuals' behaviour by limiting activities, influencing choices, or

[7] Peter Flach, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data* (Cambridge; New York: Cambridge University Press 2012), 20.
[8] "A machine learning system appears to a user or operator as composed of three elements or stages: training data, learning algorithm, and model application." *See* Matteo Pasquinelli "How a Machine Learns and Fails: A Grammar of Error for Artificial Intelligence'" (2019) 5 Spheres1, 5.
[9] Emre Bayamlıoğlu, "Contesting automated decisions: A view of transparency implications" (2018) EDPL 4(4): 433-446. https://doi.org/10.21552/edpl/2018/4/6.
[10] Yiannis Colakides, *State Machines: Reflections and Actions at the Edge of Digital Citizenship, Finance, and Art* (Institute of Network Cultures 2019), 147.
[11] What is referred to here as facts/evidence may either be simple data features (e.g., age, gender) or composite features that combine multiple dimensions of the data as inferences relating to one's education level, monthly income and etc. *See* Bayamlıoğlu (n 9).

creating new scopes for action.[12] The normative scrutinization of an automated decision is a question relating to the regulatory or legal domain that the system operates in, e.g., school admittance, distribution of social benefits, immigration applications, law enforcement, medical diagnosis, insurance, recruitment or work performance.

Transparency in ADM, as defined above, will entail that individuals should know what data were collected or were found to be relevant for a decision about them, and eventually how these data were assessed. Put in other words, the aim is to scrutinise whether the data and the way it is used provide a reliable basis to draw inferences within a given context, and whether these inferences lead to normatively acceptable decisions.[13] Evidently, such abstract and multi-layered conception of transparency may not be fully operationalised through unmediated or unfiltered human access to data or the algorithms. Transparency for the purposes of contestation rather requires a variety of *technical*, *administrative* and *procedural* measures, which equally takes into consideration both the outcome and the process itself.[14]

2.2 The relevance of IP rights and the structure of the analysis

What follows from above is that rendering ADM systems transparent for the purpose of contestation may be of relevance to IP rights in a variety of ways: *First*, transparency of ML-based decisions may involve an analysis (and thus reproduction) of both the *actual data* processed and *the training data* used to construct and calibrate the ML model.[15] *Second*, to identify systemic anomalies, the scrutiny may also extend to the analysis of the system's output (both as a database of past decisions and as individual results). For instance, by assessing the relevance of the online ads, one can form an intuitive causal explanation about how an online advertising system selects the content. *Third*, in many cases, the insights into the algorithmic techniques, their mode of deployment and real impact on a specific decision, may not be sufficiently understood without access to the computer code. *Fourth*, the development of software tools to audit ADM systems or to contest individual decisions may require the adaptation, alteration or implementation of certain parts of the ML system (e.g., for a modelling or simulation imitating the operations and processes of the system to test and discover decisional criteria).

As such, IP rights potentially contain an arsenal of legal remedies that may be relied upon by the operators and developers of ADM systems to confront transparency demands at several fronts. Many informational elements and essential components in a ML process may fall within the ambit of IP protection—being subject to copyright, *sui generis* database rights, patents or trade secrets. That is, the source code, training data, and several elements or aspects of the ML model may be proprietary, and —to the extent that they qualify as confidential business assets—can be shielded from access as trade secrets.

[12] Michael Latzer and Noemi Festic, 'A guideline for understanding and measuring algorithmic governance in everyday life' (2010) 8(2) *Internet Policy Review*.
[13] Sandra Wachter and Brent Mittelstadt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI" *Columbia Business Law Review*, *2019*(2): 494–620.
[14] Emre Bayamlıoğlu, 'The right to contest automated decisions under the GDPR: Beyond the so-called "right to explanation"**,** *Regulation & Governance* 16(4): 1058–78. https://doi.org/10.1111/rego.12391.
[15] Benjamin L. W. Sobel, "Artificial Intelligence's Fair Use Crisis". *The Columbia Journal of Law & The Arts*, (2017) 41 (1):45-97, 61.

IP rights, as *erga omnes* exclusive rights, are legal constructs which govern the use and control of information in different form and function. Consisting of several segments—aiming, by and large, to foster innovation and dissemination of knowledge— IP regime was never designed as a uniform project.[16] Until the emergence of digital technologies, each segment was treated as an isolated compartment and the interplay between different protection categories or subject-matters was mostly ignored.[17] With the pervasion of internet and other digital technologies, the boundaries of these segments expanded and began to overlap with each other. This resulted in creations, inventions or informational products capable of being protected under more than one intellectual property regime (e.g., software protected by both patent and copyright). Digital technologies make transition between different protection modalities easier and inevitable, giving rise to uncertainties which had not been anticipated at the time of the inception of the relevant IP regimes.[18] Therefore, to properly deal with the implications of IP rights as counter-arguments to transparency demands in ADM, the below structure (which will be developed in the rest of the paper) is not based on the segmentation of IP rights but rather on an ontology of ML as: *expressional* and *operational* (utilitarian/functional) elements.[19]

i) Protection on individual items (works) contained in datasets, e.g., user-generated text, image etc. (Part III - *expressional elements*).

ii) Protection relating to datasets (Part III - *expressional elements*).
- actual data under analysis,
- training data,
- output data (either as a compilation of past decisions or an individual result/decision).

iii) Protection relating to operational elements of the system (Part IV- *operational/ functional elements*).
For the purpose of IP analysis, operational elements of ML systems may be taxonomized as: *algorithmic techniques (algorithms)*, *computer code*, and the *ML model.* These further split as literary and utilitarian elements in Parts 4.2 and 4.3.

iv) Trade secret protection for all relevant ML elements (Part V-).
Trade secret protection could relate to all components and informational elements contained in ML-based decision-making systems.[20] Trade secrets either overlap with other IP rights (as an additional layer of protection) or serve as a reserve legal remedy for ML elements which are not eligible to IP protection.

[16] Martin Husovec, "The Essence of Intellectual Property Rights Under Article 17(2) of the EU Charter" (2019) *German Law Journal, 20*(6): 840-863.
[17] Robert Tomkowicz, *Intellectual Property Overlaps: Theory, Strategies and Solutions,* (New York: Routledge 2012), 1.
[18] *ibid,* 5.
[19] Part V, on trade secrets, is an exception to this structuring as analysing all elements under a single protection regime (The Trade Secrets Directive). This is for the reason that trade secret law is not confined to a specific type of intellectual labour (e.g., artistic/literary work, database or technical invention) but protect any type of information against unlawful appropriation.
[20] ML-based decision systems are technical assemblages arranged in multiple layers and comprising of heterogenous and vaguely describable elements such as algorithms, data models, data structures, system metrics, computer code and hardware.

Based on the above structure, the rest of the paper will identify the possible IP infringements and analyse to what extent reliance on IP rights could relieve the automated decision-makers from the obligation of making transparent and contestable decisions (e.g., under Article 22 of the GDPR).

3. IP protection pertinent to *data* and *datasets*

3.1 A general overview of data and datasets in ML systems

Considering the above-explained transparency requirements aiming for contestability, we can identify three categories of *data/databases* that are relevant to ADM process: i) the training data, ii) the "actual" data (a.k.a. input data[21]) and databases analysed for a specific decision, iii) data and databases comprising of ML output (predictions, ratings, etc.). This relies on a distinction between the data used during the development stage and the (actual) data used by an AI system to produce a specific result.

The knowledge of the *actual data* being processed is a precondition so that individuals can form an intuitive causal explanation about how their traits, emotions or actions make the ground for a specific decision.[22] ML systems analyse data to map an input to an output, based on a set of examples referred to as *training data*. Training data provides insights into the data features that are factored in for a specific decision. The knowledge of training data helps understand what the system learns in relation to processes, things, or the people under analysis as well as the interactions between them. Depending on the specific ML task, training data may consist of a body of case law, a collection of photographs, a database of statistics or a record of machine readings (sensor data). The third category relevant for contestation is the *output data*[23] either as an individual ML result or in the form of a database compiling

[21] The AI Regulation proposal introduces categories for ML data as *training data, validation data, testing data* and *input data*) and requires that each category complies with certain quality criteria Proposal for a Regulation of the European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final} (hereafter AI Regulation).

[22] In case of automated decisions about real persons, a substantial part of the "actual data" contains traces that render the individual identifiable and thus, falls under the definition of personal data of the EU data protection regime. In such cases, the data subject contesting an automated decision will have a direct right to access her/his personal data (Article 13 and 14 of GDPR) without recourse to transparency requirements implicit in the right to contest in Article 22.

[23] The concept of *output data* partly matches with the concept of "derived data" (aka predicted or inferred data) which is discovered by the analysis of provided or observed data —e.g., predicting residential stability, verified employment, modelled income, spending capacity, education level, racial origins or peak times in a commercial premise. Databases that contain trivial data about the individuals can be used to generate unknown data(information) about their likely identity, attributes, interests, or demographic features. Despite its importance, data subjects have little to almost no control over their derived data. The term derived, inferred or predicted data occurs nowhere in the text of the GDPR. Nor there is any explicit provision in the GDPR which obliges data controllers to disclose such type of information. This leaves it open, to what extent this broad category qualifies as personal data and which of the data subject rights provided in the GDPR apply. On this crucial question, WP29 has rendered opinion in its Guidelines on Automated Individual Decision-Making and Profiling, stating that certain individual rights (e.g., right of access, the right to rectification, erasure and restriction) may apply to derived data. However, WP29 has also made it clear, this is not without restrictions. Even if derived data is accepted as personal data, there are still practical and theoretical setbacks regarding the exercise of the data subject rights. For instance, WP29 Guidelines on Data Portability limit the scope of the right (Art. 20 GDPR) with the personal data provided by the data subject or observed by the controller, and thus exclude derived data. For more, see Gianclaudio Malgieri, "Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data" (2016) *PinG Privacy in Germany*, no. 4 :5; Paul De Hert, Vagelis Papakonstantinou, Gianclaudio Malgieri, Laurent Beslay, and Ignacio Sanchez., "The Right to

various results or past decisions.[24] Access to a large sample of results may be necessary to reveal how different data inputs changed the outcome, and consequently to prove any objectionable conduct.[25] Having access to a collection of ML output may enable the reverse-engineering of both the decisional criteria and the inferential logic employed by the system.

For all three categories (the training data, "actual" data and output data), both the databases and the individual data items may enjoy the protection of copyright so long as they qualify as creative expressions (below 3.2). The second type of protection that will be analysed in this Part, the *sui generis* database right, is only applicable to databases (below Part 3.3).

3.2 Copyright as artistic or scientific expression

In ML-based decision-making systems, copyright protection of *data* and *datasets* as artistic or scientific expression may be of relevance in two ways. *First*, the individual items in a database may be eligible to copyright protection in their own right (3.2.1). *Second*, databases (as compilations) may enjoy protection due to the creativity in the selection and the arrangement of the items comprising them (3.2.2).

3.2.1 Copyright protection of individual items (works) contained in the datasets

While the human-created elements contained in a data corpus such as text, images, videos, sound recordings may be subject to copyright, factual information (data) either provided by the user or captured through sensing or tracking are not copyright-eligible as they lack originality. Under the EU law, this is laid out in the InfoSoc Directive and applies both to training and actual data used in the development or deployment of ML systems. The copyright on human-created elements may be owned by several different actors, including the individual who is subject to an automated decision.[26]

The other data category, ML output, may take the form of a classification, numeric score, binary decision, or a textual suggestion. ML output not only covers certain (empirical) verifiable information (e.g., spending capacity) but also subjective evaluations, predictions, opinions, or assessments such as one's possible age of death, future professional misconduct or emotional state. In terms of copyright eligibility, a distinction could be made between the verifiable empirical findings and the subjective inferences about behaviours, events or risks. The former may be seen as statements of facts which are either true or false, and as such they are indisputably out of the scope of copyright protection. Coming

Data Portability in the GDPR: Towards User-Centric Interoperability of Digital Services'. (2018) *Computer Law & Security Review* 34(2):193–203.

[24] In German OpenSCHUFA project, 2800 volunteers provided their personal credit report to identify systemic anomalies in the scoring system. Findings indicate that Schufa certifies a higher risk to young men and former low scores of an individual might adversely affect the results. See 'OpenSCHUFA – shedding light on Germany's opaque credit scoring' (*Algorithm Watch*, 21 February 2018) <www.algorithmwatch.org/en/openschufa-shedding-light-on-germanys-opaque-credit-scoring/>

[25] Frank A. Pasquale, "Restoring Transparency to Automated Authority" (2011) *Journal on Telecommunications and High Technology Law*, 9: 235-254.

[26] When created by the users of online services, such material is referred to as User-generated Content (UGC) which is generally a derivative of the existing works possibly subject to third party copyright. Regarding copyright matters on UGC, *see* Daniel J. Gervais, "The Tangled Web of UGC: Making Copyright Sense of User-Generated Content" (2009) *Vanderbilt Journal of Entertainment and Technology Law* 11(4): 841-870. Given that users generally have no economic expectation over the content they have created during their internet use, it may be argued that utilisation of this content for the purposes of scrutiny will not harm the legitimate interests of the rightholders and therefore may be regarded to be of *de-minimis* nature.

to subjective inferences, although they may be regarded as creative opinions, they generally do not amount to a copyrightable form. Copyright protects expressive forms and does not extend to information contained in the work— however novel or original the information might be. The outcome of a ML model, be it predictions or credit scores, is not copyrightable as being merely abstract information.[27] Creativity counts for copyrightability only if it translates into a concrete original expression.[28] For instance, an original graphic interface for representing or visualising ratings may be subject to copyright protection (e.g., something more original and creative than "☆☆☆☆☆" to rate hotels, restaurants and etc.). By the same token, while a personal profile as abstract information is not copyrightable, the literary, visual or audial expression of it in a tangible form may be eligible to protection.

Irrespective of the fact that AI systems are increasingly capable of taking over human-specific tasks (including the creation of literary texts, melodies or images), ML outcome could be excluded from copyright also due to the unrecognised authorship status of the machines. Despite the lack of an explicit reference, both the EU copyright regime and Berne Convention[29] require human interference as a condition for copyright protection.[30] Having said that, copyright law is not totally alien or anathema to the idea of quasi-authorship status of entities other than natural persons. For example, neighbouring or derivative rights (as a special extension of copyright regime) exclude moral rights for phonogram and film producers and provide a fixed term of protection starting from the date of release. Moreover, as some common law jurisdictions are more permissible in terms of assigning authorship to legal entities, it seems difficult to apply the human interference requirement strictly as a universal principle.[31]

3.2.2 Copyright in databases based on the selection and arrangement of the individual data items

Even though factual data itself is out of the scope of copyright protection, the databases (compilations of data) which present creativity as to their selection or arrangement could be eligible copyright

[27] What is referred to here as "output data" is also information in machine-usable form. *See* Sasa Baskarada and Andy Koronios, 'Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of The Hierarchy and its Quality Dimension' (2013) 18 Australasian Journal of Information Systems 5.

[28] James Grimmelmann, "Three Theories of Copyright in Ratings" (2012) 14 Vanderbilt Journal of Entertainment and Technology Law 851, 878; Daniel Schönberger, "Deep Copyright: Up-and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML)" in Jacques De Werra (ed), *Droit d'auteur 4.0 / Copyright 4.0* (Schulthess Editions Romandes 2018), 145.

[29] Berne Convention for the Protection of Literary and Artistic Works (Paris Act of 24 July 1971 as amended on 28 September 1979).

[30] Accordingly, the term (duration) of protection which is linked to the author's lifetime and the catalogue of moral rights only makes sense in case of a human creator. Sam Ricketson, 'The 1992 Horace S. Manges Lecture -People or Machines: The Berne Convention and the Changing Concept of Authorship' (1992) *The Columbia Journal of Law & the Arts* (1991-1992) 16(1).

[31] It is also possible that the copyright in computer-generated works could be allocated to the user of the generator program. Section 9(3) of the U.K. Copyright, Design and Patents Act (1998) provides that in case of computer-generated works, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken. For more, see Pamela Samuelson, "Allocating Ownership Rights in Computer-Generated Works" (1986) 47 *University of Pittsburgh Law Review*, 1185-1228, 1186; Stef van Gompel, "Creativity, autonomy and personal touch: A critical appraisal of the CJEU's originality test for copyright" in Mireille van Eechoud (ed), *The Work of Authorship* (Amsterdam: Amsterdam University Press 2014), 95-144; Jean-Marc Deltorn, "Deep Creations: Intellectual Property and the Automata'" (2017) 4(3) Front. Digit. Humanit.

protection.[32] All databases pertinent to ML process (except for restrictions about the machine authorship) can benefit from this protection— irrespective of whether or not the individual data items contained in the datasets are themselves creative expressions subject to copyright protection.

Under the EU Database Directive[33], databases may be eligible to copyright protection as creative expression, where selection and arrangement of the content presents originality (a stamp of the human creator).[34] Article 3(1) of the Directive provides that "databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation" are protected by copyright. Copyright on databases applies only after a selection or arrangement of data and thus, the creativity requirement sets a limit which prevents the expansion of protection to raw data aggregations.[35] This protection provides a lesser catalogue of rights (in comparison to InfoSoc Directive), only prohibiting temporary and permanent reproduction, alteration and further dissemination of the proceeds of the alteration. Needless to say, the protection does not extend to the contents of the database.

In a 2012 judgment, Court of Justice of the European Union (CJEU) ruled that originality should be understood in the sense that the author, through the selection or arrangement of the data, expresses her creative ability in an original manner by making free choices.[36] That is, labour and skill alone is not sufficient. For instance, the construction of a training dataset may be laborious and may require a certain amount of creative thinking and deliberate choices on the side of the programmer. However, in order to be eligible to copyright, the creativity must be reflected in the resulting work in a way sensible by human cognition.[37] Although neither the training nor the actual data may be regarded as solely dictated by technical constraints or operational needs, drawing a clear line between the "utilitarian" and "expressive" aspects a database (in terms of its organization or the structure) is not an easy task. Since ML databases are constructed for practical ends in a standardized and automatized manner, there is hardly any dimension which might be considered to be independent of the purposes that the system has been built for—leaving very limited room for personal imprint.[38] Regarding the databases comprising of ML output, they could be held ineligible to copyright-protection for reasons mentioned above (3.2.1) about the machine authorship as well as the explicit reference to natural

[32] Art. 10(2) of the Agreement on Trade-Related Aspects of Intellectual Property Rights, adopted in Marrakesh on 15 April 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C (TRIPs) and Art. 5 of the World Intellectual Property Office Copyright Treaty, adopted on Dec. 20, 1996, WIPO, Doc. CRNRIDC/94 (WCT).
[33] Council Directive (EC) 96/9 of 11 March 1996 on the legal protection of databases [1996] OJ L 77 ('Database Directive').
[34] Mark J. Davison, *Legal Protection of Databases* (Cambridge: Cambridge University Press 2003)
[35] Francesco Banterle, "Data Ownership in the Data Economy: A European Dilemma" in Synodinu, Tatianē-Elenē, Philippe Jougleux, Christiana Markou, and Thalia Prastitou, eds. *EU Internet Law in the Digital Era: Regulation and Enforcement*. (Cham, Switzerland: Springer 2020), 199-225.
[36] Case C-604/10 *Football Dataco Ltd and Others v Yahoo! UK Ltd and Other*s [2012] EU:C:2012: 115.
[37] See above (3.2.1) explanations about *output data*.
[38] Andreas Wiebe, "Protection of Industrial Data – a New Property Right for the Digital Economy?' *Journal of Intellectual Property Law & Practice* 12(1): 62–71; Banterle, (n 35), 207. For a more permissive approach to the problem, see Inge Graef, "Market Definition and Market Power in Data: The Case of Online Platforms" (2015) *World Competition* 38(4) :473–505. For earlier US judgments deciding that developing programs for interoperable data formats was not infringement and the structure and organisation of computer input formats could not be protected for being determined in large part by functional considerations, see Rosa Maria Ballardini, "Scope of IP Protection for the Functional Elements of Software" in *In Search of New IP Regimes* (IPR University Center 2010), 40, fn 38.

persons in Article 4 of the Database Directive. Yet, the cases where the output data is indirectly influenced by the designer or the operator of the system remain open to debate.

3.2.3 Exceptions to copyright protection as artistic or scientific expression

Copyright law neither aims to conceal nor to prevent the use of information, but it only protects the original expression of the author. In the abstract, transparency demands, so long as limited to access to information, do not give rise to a copyright infringement. Yet, the deployment of transparency mechanisms involving audit and black-box testing may at some stage infringe the exclusive rights (in particular reproduction and adaptation) of the copyright holder.

Both in international treaties (Berne Convention, 1996 WIPO Copyright Treaty, TRIPS Agreement) and in the EU Infosoc Directive, the right of reproduction is understood in a broad sense to include every act of "copying" either in digital or physical form irrespective of its economic or functional significance.[39] In addition to reproduction, the transparency measures may also trigger the right of adaptation (Art. 12 Berne Convention) as they require switching between data formats or selecting certain data from the rest of the corpus.[40] Although right of adaptation is not covered by the InfoSoc Directive and thus not harmonised at the EU level, it is generally regarded to be implicit in the right of reproduction.[41]

Regarding the reproduction of individual data items, the compulsory exception allowing acts of temporary reproduction provided by Article 5(1) of the InfoSoc Directive is subject to narrow interpretation. To benefit from this exception a temporary copy should be transient or incidental as an integral and essential part of a technological process— aiming solely either enabling a transmission in a network between third parties by an intermediary, or a *lawful use*[42] of a work or protected subject-matter. Copying, porting and further use of the datasets for the purpose of scrutiny may not be easily kept within the confines of the provision in that the scrutiny may require more than a purely transient copy. Hence, the possible applicability of this exception might differ according to the data usage necessary for a particular transparency measure.

Software tools operating within the limits of this exception such as completing scrutiny without any permanent reproduction or alteration may be seen as an important part of the efforts in the implementation of the right to contest. However, as will be seen below (3.3.2), such exception is not available for sui generis database protection.[43]

---

[39] Recital 21 of the InfoSoc Directive.

[40] In principle, other than right of reproduction and adaptation, the audit and testing of ADM systems do not infringe exclusive rights such as *distribution* or *making available to public*. More on this, *see* Jean-Paul Triaille, Jérôme de Meeûs D'Argenteuil and Amélie de Francquen, *Study on the Legal Framework of Text and Data Mining (TDM),* European Commission, Directorate-General for the Internal Market and Services, Publications Office. (2014), 31.

[41] In contrast, under Article 5(a) of the Database Directive, the copyright protection regarding the arrangement and selection of data items explicitly includes translation, adaptation, arrangement and any other alteration. Christophe Geiger and and Giancarlo Frosio, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects" (2018) CEIPI Research Paper 2018-02, fn.12. < http://ssrn.com/abstract_id=3160586>

[42] For the meaning and interpretation of the term lawful/user in EU copyright acquis, see below 3.3.2 and 3.4.

[43] Software tools operating within the limits of this exception such as completing scrutiny without any permanent reproduction or alteration may be seen as an important part of the efforts in the implementation of the

Another compulsory exception which may be of relevance is Article 6(1) of the Database Directive regarding the selection and arrangement of the contents of a database. The exception permits restricted acts where they are necessary for the purposes of access to the contents of the databases and for the normal use of the contents by a lawful user.[44] In the usual interpretation of the provision, the acts for scrutiny or contestation cannot be easily regarded as a "normal use" of the database. Nevertheless, copyright in the selection and arrangement of the databases does not seem to pose a significant barrier. In comparison to exclusive rights on artistic and scientific works, it is a "thin" protection which applies only if the expressive and creative aspects of the database are distinctly appropriated.[45] The extraction of information from a database is not an infringement since abstract information does not relate to the creative dimension of the database.[46]

3.3 *Sui generis* database right

3.3.1 The scope of the sui generis right: eligible databases in ML context

The EU Database Directive provides also for a s*ui generis* database right as a special type of protection recognized on the entirety of a database. The *sui generis* right only protects the database as a collection and does not extend to individual data— thus keeping semantic information contained in the database in the public domain.[47] This protection, unique to the EU, is without regard to any creativity either as to the content or to the selection or arrangement of the database.[48]

Article 1 of the Directive provides a broad definition of a *database* as a collection of independent works, data or other material arranged in a systematic or methodical way, and individually accessible by electronic or other means.[49] Irrespective of their copyright eligibility, databases protected by *sui generis* right may consist of any sort of material, in any form whether electronic, in paper, online, or hybrid. The Directive does not intend to grant property rights on individual data items but protects the investment in the database against the extraction or the re-utilisation of its contents.[50]

---

right to contest. However, as will be seen below (3.3.2), such exception is not available for sui generis database protection.

[44] "This exception was inspired from a corresponding provision in Article 5(1) of the Software Directive." See Triaille and others (n 40), 72. On the concepts of "lawful user", "lawful use" and "lawful access" in EU copyright law, see below 3.3.2.

[45] As Hugenholtz states "copying a substantial part of the data without appropriating, either in whole or in part, the selection or arrangement of the data, […] will not amount to copyright infringement, but most likely will infringe the sui generis right." P. Bernt Hugenholtz, "Something Completely Different: Europe's Sui Generis Database Right" in Susy Frankel and Daniel Gervais (eds), *The Internet and the Emerging Importance of New Forms of Intellectual Property, (*the Netherlands*:* Kluwer Law International 2016), 205-222, 216.

[46] Josef Drexl, *Data Access and Control in the Era of Connected Devices,* Study on Behalf of the European Consumer Organisation BEUC (2018), 86. For more, *see* Marcelo Corrales Compagnucci*, Big Data, Databases and Ownership Rights In The Cloud* (Singapore Springer 2020), 23-24.

[47] Josef Drexl, 'Designing Competitive Markets for Industrial Data' *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, (2016) 8(4) 257 para 1, 269.

[48] For more on the Directive, see Hugenholtz, (n.45).

[49] For an interpretation of the definition of "database", see Case C-490/14 *Freistaat Bayern v Verlag Esterbauer* [2015] EU:C:2015:735.

[50] Francesco Banterle, "The Interface between Data Protection and IP Law. The Case of Trade Secrets and the Database *sui generis* Right in Marketing Operations, and the Ownership of Raw Data in Big Data Analysis" in Bakhoum, Mor, Beatriz Conde Gallego, Mark-Oliver Mackenrodt, and Gintarė Surblytė-Namavičienė (eds), *Personal Data in Competition, Consumer Protection and Intellectual Property Law: Towards a Holistic Approach?* (Berlin, Heidelberg: Springer Berlin Heidelberg), 411–443, 435.

Extraction in Article 7(2a), as a broad concept, refers to permanent or temporary and direct or indirect transfer of contents by any means or form—including the mere act of accessing and reading the database.[51] In ML context, uses of the data for the purpose of auditing or testing of the ADM system might involve the extraction of all or a substantial part of the data held in the database. This also applies to indirect or incremental ways which lead to the reconstitution of at least a substantial part of the database. Re-utilization, on the other hand, as defined in Article 7(2b) is normally not relevant in the context of contestation as it deals with the dissemination of the database by way of distribution of copies, or other forms of transmission.

Although Recital 45 makes it clear that the *sui generis* right does not constitute an extension of the protection to mere *facts* or *data*, it is subject to debate what type of processing and structuring render raw datasets eligible to *sui generis* protection. In ML applications, especially the training datasets usually go through intense pre-processing and transformation, thus qualify for *sui generis* protection as an organized set. However, where the system uses unstructured data such as books or pieces of music or video, such corpus may be excluded from protection for not being systematically organised.

Protection under the Directive requires a qualitatively and/or quantitatively substantial investment in either *obtaining, verification* or *presentation* of the contents of the database. Under this requirement, datasets comprising of output data may not be eligible to protection based on the so-called *spin-off* doctrine established by the European Court of Justice in a series of judgments in 2004.[52] According to the Court, where the 'creation' of data and the subsequent database is a by-product of the database maker's main activity, and the investment is confined to this main activity (and not to the collection of existing data), such database shall not be protected by the *sui generis* right. In this interpretation, the investment refers to the resources used to incorporate existing independent material.[53] The court has later limited the application of the doctrine in its *Football Dataco* judgment by holding that facts collected about a football game such as the score, scorer, or penalty decisions were not 'created' data. Yet, under the doctrine, it remains open whether the databases used by ML systems can always satisfy the substantial investment requirement regarding the obtaining, presentation or verification of the machine-generated output.[54]

According to Hugenholtz, it is not clear in the CJEU judgments how the machine-generated data could be situated within the spectrum between the purely *synthetic data* and the *data observed.* As he puts: "The answer depends on the type of data that the machine processes. For example, sensor data produced by a radar system or observation satellite are likely to qualify as data 'observed'. Conversely, computer-

[51] For more on the term of "extraction", *see* the Case C-304/07 *Directmedia GmbH v Albert-Ludwig Universitat Freiburg* [2008] EU:C:2008:552. Also *see* Compagnucci (n 46).
[52] Case C-46/02 *Fixtures Marketing Ltd v. Oy Veikkaus Ab* [2004] ECR I-10365; Case C-203/02 *The British Horseracing Board Ltd and Others v. William Hill Organization Ltd* [2004] ECR I-10415; Case C-338/02 *Fixtures Marketing Ltd v. Svenska Spel AB* [2004] ECR I-10497 ; Case C-444/02 *Fixtures Marketing Ltd v. Organismos prognostikon agonon podosfairou AE (OPAP)* [2004] ECR I-10549.
[53] The Court found that the databases such as a football fixture or a horse race bulletin did not deserve protection under the Directive as the they were the by-products of the main activities, namely managing horseracing or running the football league.
[54] Unlike, ML output compiled into a database such as search engine results, not every ML-based system creates outcome which is accessible in the form of a readily available database generated by the system. For instance, in an online advertising system, consumers/users are not provided with a selection of results, but simply exposed to a certain content. In that case, the analysis of different advertising messages for the purposes auditing the system (e.g, to discover the data features that affect the price of the advertised product or service), cannot possibly infringe sui generis right simply because *extraction* and *reutilisation* primarily require a pre-existing database.

generated airline schedule data squarely falls under the rubric of 'created' data excluded by the European Court."[55] Accordingly, the Court's epistemological distinction between "creating" and "obtaining" data is not self-evident, and, in any case, the protection of machine-generated data brings *sui generis* protection closer to a "property right". Some commentators do not find this conclusion warranted, as it severely limits the application of the Directive, for instance, in the Internet of Things (IoT) environment.[56] Leistner argues that, until the CJEU further clarifies the matter, drawing the line between obtaining and creation of data would remain contentious and legal uncertainty would prevail.[57] In sum, it is difficult to argue that the choice of the terms *obtaining, verification* or *presentation* at the time of the enactment of the Directive was intended to exclude machine-generated data. The emphasis of the Directive (in Recitals 45 and 46) that *sui generis* database right was not an extension of copyright protection to mere facts or data primarily reflects the concerns about the monopolization of the semantic content of the data.[58] Therefore, denying protection in cases where the maker of the database is the sole holder of the information contained in the data could better serve to the purposes and the rationale underlying the spin-off doctrine.[59]

3.3.2 Exceptions to sui generis database right

To begin with, the Database Directive does not provide a temporary reproduction exception similar to Article 5 of the InfoSoc Directive. Nevertheless, the Directive provides a number of exceptions in Articles 8 and 9 which may be relevant for the purpose of contesting automated decisions.

Article 8(1) provides that the *lawful* user of a database, which is made available to the public, could extract and/or re-utilize the insubstantial parts of its contents. As Article 15 of the Directive declares any contractual provision contrary to Articles 6 (1) and 8 as null and void, the extraction of the insubstantial parts of a database may not be prohibited through user agreements or license contracts.[60]

---

[55] Hugenholtz also mentions that when concomitant investments are taken into account, the output data may be regarded to involve substantial investment independent of the resources allocated to create the data. P. Bernt Hugenholtz, "Against 'Data Property" in Hanns Ullrich, Peter Drahos and Gustavo Ghidini (eds), *Kritika: Essays on Intellectual Property,* (Cheltenham, UK: Edward Elgar Publishing Limited 2018). Also *see* Estelle Derclaye, "The Database Directive" in Irini Stamatoudi and Paul Torremans (eds), *EU Copyright Law* (Cheltenham, UK ; Northampton, MA: Edward Elgar Publishing, 2021), 216–254.

[56] Graef argues that spin-off will not be applicable to the "inferred data" accumulating in the hands of the online platforms. Graef (n 38) 484.

[57] Matthias Leistner, "Big Data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform" in Sebastian Lohsse, Reiner Schulze, and Dirk Staudenmayer (eds). *Trading Data in the Digital Economy: Legal Concepts and Tools*. (Oxford, England]: Nomos ; Hart Publishing, 2017).

[58] "The contents of a database are information in the widest sense of that term." See Banterle, (n 50) 423. Also *see* P. Bernt Hugenholtz, 'Implementing the European Database Directive' in Jan J.C. Kabel and Gerard J.H.M. Mom (eds), *Intellectual Property and Information Law, Essays In Honour Of Herman Cohen Jehoram* (The Hague; Boston: Kluwer Law International, 1998), 183-200.

[59] European Commission's proposed regulation Data Act does not make any distinction between obtained or created data while exempting IoT data from protection. See below 3.5 and Margoni (n 88).

[60] In the case *Ryanair*, the CJEU ruled that for databases which do not qualify for *sui generis* protection, Directive's Articles 6(1), 8 and 15 (which preclude contractual limitations) did not apply and thus, parties were free to determine the conditions of use through a contract. The judgment gave rise to a paradoxical situation in that the databases out of the scope of the Directive received stronger protection through contracts. As presuming an *ab initio* "right" on data, the Court's approach was criticized for contradicting with the rationale of the Database Directive. *See* Maurizio Borghi and Stavroula Karapapa, "Contractual restrictions on lawful use of information: sole-source databases protected by the back door?" *European Intellectual Property Review*, (2015), 37(8):505-514.

The lack of clarity, as to the extent that the use of the database will be regarded as substantial, may be a significant drawback regarding the applicability of the provision. More importantly, it should be noted that Article 8(1) and in particular the reference to *lawful user* have been heavily criticised for being somewhat redundant. As the Directive protects only substantial part of databases, this implies that the maker of the database has no exclusive rights on the insubstantial parts which need to be restricted.[61] It is argued that the reference to a lawful user in Article 8(1) may give rise to confusion and even an interpretation of the provision expanding the *sui generis* right into insubstantial parts of databases.[62]

The provision, limited to lawful users, is further curtailed by Article 7(5) which prohibits repeated and systematic extractions of a database aiming at reconstituting the whole or a substantial part of the contents of a database. Accordingly, where numerous data subjects collectively demand copies of their profiles (e.g., via a *data trust* or *data cooperative* representing them), it is open to debate whether such demand could be objected on the ground that the cumulative effect of the bulk request would amount to substantial extraction. Nevertheless, in the BHB case, the CJEU ruled that Article 7(5) would be applicable only if the cumulative effect of the repeated acts seriously prejudice the investment in a manner amounting to an extraction and/or re-utilisation as referred to in Article 7(1) of the Directive.[63] Hence, a scrutiny analysis may be carried through repeated and systematic extractions of insubstantial parts so long as the purpose is not to reconstitute the whole or a substantial part of the database.[64]

Article 9 of the Database Directive further provides three non-mandatory exceptions for the benefit of the lawful user of a database which is made available to the public. Article 9(1b) allows for extracting or re-utilizing a substantial part of the contents of a database for the purposes of illustration for teaching or scientific research to the extent justified by the non-commercial purpose. What is more noteworthy is Article 9(1c) which permits lawful users to extract and/or re-utilize the contents of a database for the purposes of public security, or an administrative or judicial procedure. This is the most relevant exception in the Database Directive which could be implemented to give effect to the right to contest automated decisions as provided under the GDPR (Art.22). Given that failure to provide the necessary means to contest an automated decision under Article 22 may result in legal proceedings against the data controller, transparency requirements may be kept exempt from the *sui generis* protection under this provision. Lastly, Article 9(1a) provides a private use exception which, under certain conditions, could facilitate transparency efforts.

The notion of *lawful use/user*, provided as a condition to benefit from the exceptions in Database Directive (1996), first appeared in EU copyright *acquis* by the Software Directive (1991).[65] Both Directives employ the concept in a similar fashion to legally guarantee a minimum space of free use

[61] Triaille and others assert that, unless "insubstantial "is understood *somehow* in a totally different context, the provision is" illogical" in that the use of insubstantial parts do not fall within the scope of the extraction right and thus, the maker of the database has no exclusive right to restrict such acts. Triaille and others (n 40) 76.
[62] On this matter, see Cristina Angelopoulos, 'Database Directive' in Thomas Dreier and PB Hugenholtz (eds), *Concise European copyright law* (Second edition, the Netherlands:Kluwer Law International 2016), 409.
[63] C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd.* [2004] ECR I-10415, para.85.
[64] Triaille and others (n 40) 79.
[65] Council Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Codified version) [2009] OJ L 111/16. In order to determine the person who can lawfully invoke the application exceptions to the copyright protection on software, the Software Directive uses terms such as *lawful acquirer of the program* or *person having a right to use the computer program* indiscriminately. See below 4.2.1.

against restrictive contractual arrangements of the rightholders.[66] The need for such allowance stems from the fact that both types of protection (software protection and sui generis right), to some extent, give control over the use of information whereas conventional copyright only regulates the commercial exploitation of artistic and scientific expression. The concept of lawful use is also a condition of the temporary copy exception of the InfoSoc Directive (in Article 5(1)) which permits the making of electronic copies as part of a *lawful* or *authorised* use.[67] The CJEU has interpreted the term in several judgments within the context of Article 5(1).[68] These judgments do not define lawful user within the meaning of Database Directive but rather enumerates permitted types of reproduction of or access to works.

Under the Database Directive lawful user could be understood as the person who acquires the database or a licit copy of it in a lawful way. "Broadly speaking, a lawful user would be a person who has obtained the copy of the work or the right to use the work, without infringing copyright laws."[69] This may be through a subscription agreement providing access to a restricted content or by way of a statutory exception. For databases accessible without payment or password, such as websites with unrestricted access, an implied (license) contract could be contemplated within the limits of the Directive. Leaving this aside, in majority of the cases an individual who is subject to an automated decision would fail to qualify as the lawful user of the database.

3.4 EU Directive on copyright in the Digital Single Market and the text and data mining (TDM) exception

The Directive, *Copyright in the Digital Single Market* (DSM Directive)[70], aiming to reform the EU copyright law, introduces two mandatory restrictions on copyright and sui generis right for the purpose of *text and data mining* (TDM) in Articles 3 and 4 of the Directive.[71] The Directive defines TDM in a way to include a great variety ML-based analytics. The exceptions provided for TDM are without prejudice to the existing exceptions and limitations explained above. Recital 9 further clarifies that the analysis of mere facts or data that are not protected by copyright do not need authorisation and thus do not require an exception. [72]

---

[66] Tatiana-Eleni Synodinou, "Who Is a Lawful User in European Copyright Law? From a Variable Geometry to a Taxonomy of Lawful Use" in Tatiana Synodinou Philippe Jougleux, Christiana Markou, and Thalia Prastitou (eds), *EU Internet Law in the Digital Era* (Cham, Switzerland: Springer 2020), 27-60, 29-30.

[67] See above 3.2.2. In addition, Article 6(4) of the InfoSoc Directive uses the term "legal access". Lawful use/access is also a requirement to benefit from the *text and data mining exception* explained in the below Part 3.4.

[68] Case C-302/10 *Infopaq International A/S v Danske Dagblades Forening* [2012] ECR I-6569; Joined Cases C-403/08 and C-429/08 *Football Association Premier League Ltd and Others v QC Leisure and Others* and *Karen Murphy v Media Protection Services Ltd* [2011] ECR I-9083.

[69] Synodinou, (n 66) 39. The notion of 'lawful user' has been criticised for its lack of clarity, potential redundancy and inherent limitations. These drawbacks and the ambiguities around the concept of 'lawful user' are regarded as barriers in terms of revealing the full potential of the EU database regime. *See* European Commission, 'Evaluation of Directive 96/9/EC on the Legal Protection of Databases' (Commission Staff Working Document) SWD(2018) 147 final, 20.

[70] Council Directive 2019/790/EC of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Council Directives (EC) 96/9 and 2001/29 [2019] OJ L 130.

[71] For similar exceptions in German, French, UK and Japanese laws, *see* Daniel Gervais, "Exploring the Interfaces Between Big Data and Intellectual Property Law" *Journal of Intellectual Property, Information Technology and Electronic Commerce Law,* 2019, 10(3), para. 38-40.

[72] "The exception is justified for three reasons. First, it transfers a core principle of copyright into the digital

Considering the unharmonized state of the research and education related exceptions of the EU copyright regime, which do not fit well with the emerging ML practices, the first exception, Article 3, is intended as a reassessment and consolidation of the existing exceptions and limitations relating to research and education activities. Article 3 is a limited exception covering only i) the reproduction of works[73]; ii) temporary or permanent reproduction relating to copyright in the selection and arrangement of databases[74]; iii) extraction and re-utilization of the databases protected by the *sui generis* right[75]; and iv) the press publishers' right.[76] Accordingly, research organisations and cultural heritage institutions have an exception for TDM of works or other subject-matter to which they have lawful access for the purposes of scientific research. "Lawful access" [77] should be understood as the availability of data through open-data policies or contractual arrangements between the rightsholders and the research organizations such as subscriptions or through other lawful means.[78] By way of an explicit reference to the relevant provisions of the InfoSoc Directive, member states are required to observe the *three-step-test* in their implementation of the exception.[79] Article 7 of the DSM Directive prohibits any contractual provision contrary to the exceptions provided in Article 3.

The second restriction, Article 4 of the Directive, allows TDM for any purpose. In comparison to Article 3, Article 4 also covers the permanent or temporary reproduction of computer programs where the normal use of the program necessitates such reproduction, and further translation, adaptation, arrangement or any other alteration of the program.[80] This broad scope of the general TDM exception provided in Article 4 for *lawfully accessible* works and other subject matter seems promising in that it could eliminate some of the copyright infringements that may arise in relation to the exercise of the right to contest. However, the provision only applies to certain exclusive rights in a fragmented way, far from being a general non-infringement exception.[81] Moreover, Article 4 is not applicable if the rightsholders bring reservations in an "appropriate manner" to restrict TDM. In cases where the content is made publicly available online, only machine-readable means are considered appropriate (Recital 18). For other content, contractual arrangements and unilateral declarations could also prohibit

---

era. Non-fictional information remains in the public domain. Second, it serves the strong public interest to encourage the generation of new knowledge which would otherwise not exist due to prohibitive transaction costs." Benjamin Raue, "Free Flow of Data? The Friction Between the Commission's European Data Economy Initiative and the Proposed Directive on Copyright in the Digital Single Market'. *IIC - International Review of Intellectual Property and Competition Law* 49(4) (May 2018): 379–383, 381.

[73] Art. 2 of the InfoSoc Directive.

[74] Art. 5(a) of the Database Directive.

[75] Art. 7(1) of the Database Directive.

[76] Art. 15(1) of the DSM Directive. The provision obliges news aggregators who link to publishers' content or use snippets, e.g., Google news, to obtain a license. The provision partially conflicts with Article 10(1) of the Berne Convention which requires member states to permit free press summaries.

[77] In the DSM Directive, European legislator preferred the terms *lawful access* and *lawfully accessible* to formulate the condition of *lawful use.* Also see above 3.3.2.

[78] For a critique of this perplexing and diverse terminology of the EU copyright acquis, see Synodinou, (n 66).

[79] Art. 7(2) of the DSM Directive reads as: "Article 5(5) of Directive 2001/29/EC shall apply to the exceptions and limitations provided for under this Title." For three-step-test see below 4.3.1.

[80] The reason for this discrepancy between Art. 3 and 4, leaving computer programs out of the scope of the scientific research exception, is unclear. Yet, of note, the inclusion of computer programs within the TDM exception resonates with the perspective laid out in this paper that the data analysis may require the implementation of the embedding computer code.

[81] Gervais, (n 71) para. 44-45. Also *see* Irini A. Stamatoudi, "Text and Data Mining" in Irini A. Stamatoudi (ed.), *New Developments in EU and International Copyright Law (* Leiden, Netherlands: Kluwer Law International, 2016), 251-282, 266.

TDM. This broad allowance for the contractual circumvention is the major shortcoming of the provision which is likely to render it inefficient.

An unanswered question regarding the TDM exceptions, which is of practical importance, is the technological protection measures (e.g., DRM) preventing access to works in the digital environment. Considering that Article 3(3) of the DSM expressly mentions that the rightsholders shall be allowed to apply measures to ensure the security and integrity of networks, it is highly questionable whether the circumvention of technological protection measures (TPMs) could be permissible in the context of TDM. Considering Article 4, as rightsholders can prohibit TDM, the deployment of a TPM by the rightsholder may be interpreted as a reservation in an appropriate machine-readable manner under the provision.[82]

3.5 The EU data strategy, and the effect of the upcoming legislation on IP protection of data

With a view to address the current shortcomings of the EU *acquis* and thus to extract full economic value of data, in its 2020 Data Strategy, the European Commission has identified several critical issues that need to be overcome to foster the availability of data, ensure data interoperability and empower individuals to exercise their rights.[83] The action plan set forth by the Strategy document includes two legislative proposals which are of significance in terms of application of IP rights to data and databases.

The proposal, Data Governance Act (DGA)[84], expands the scope of public sector data initially laid down by the Open Data and PSI Directive.[85] The DGA offers promising improvements in terms of access to public sector data subject to rights of others, i.e., personal data protection, intellectual property, trade secret protection or other commercially sensitive information. As this will open a significant part of data held by public institutions/bodies to private use, the emerging regulatory landscape relating to access to public sector data is also facilitative in terms of implementing the transparency mandates prescribed by law. Yet, the provisions provided by DGA do not lay out the mechanisms or the legal solutions how data subject to others' rights can be made available for businesses. The proposed Regulation does not interfere with the existing rights but provide for a set of harmonized basic conditions which improve access to public sector data (e.g., the requirement of

[82] It should nevertheless be mentioned that according to Recital 16, TPMs should not exceed what is necessary to pursue the objective of ensuring the security and integrity of the system and should not undermine the effective application of the TDM exception (also see Recital 14). The question, to what extent TPMs could be rendered ineffective for the sake of benefitting from statutory exceptions, has been an ongoing debate since the enactment of the InfoSoc Directive. See, Geiger and Frosio (n 41).

[83] European Commission, "A European Strategy for Data" (Communication) COM(2020) 66 Final. (European Strategy for Data). Initiatives aiming for sharing, portability and access to data have been a part of the European Commission's legislative agenda since 2015. *See* European Commission, "Digital Single Market strategy for Europe" COM (2015) 192 final. For an account of the development of this agenda, see Sebastian Lohsse Reiner Schulze, Dirk Staudenmayer, "Trading Data in the Digital Economy: Legal Concepts and Tools" in Sebastian Lohsse and others (eds), in (n 57), 13-24.

[84] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)' COM(2020) 767 Final ('Data Governance Act' or 'DGA'). The Proposal focuses on four intervention areas, namely (a) mechanisms for the enhanced use of public sector data that cannot be available as open data, (b) a certification or labelling framework for data intermediaries, (c) measures facilitating data altruism, and (d) mechanisms to coordinate and steer horizontal aspects of governance in the form of an EU-level structure.

[85] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information [2019] OJ L 172/56 ('Open Data and PSI Directive').

non-exclusivity).[86] Public sector bodies, who are holders of *the sui generis* right provided in the Database Directive, are required to exercise that right in a way which would not conflict with the re-use of data provided by the proposed Regulation. Public sector bodies allowing this type of re-use are required to be technically equipped to ensure that the rights of others are fully preserved. As such, under the upcoming DGA, we may expect the development of data analysis techniques that do not give rise to permanent copies of data or other protected elements.

The other important legislative initiative under the 2020 Data Strategy is the Data Act which aims to encourage and enable B2B and B2G data use in all sectors. The proposal which was released on 23.02.2022 introduces several interventions to the existing regulatory and contractual framework relating to data reuse and sharing both in the B2B and B2G contexts. Chapter II of the proposal provides for an obligation on the *data holders* to make available to the user (or to a third party designated by the user) the data generated by the use of the IoT product or a related service where such data are not already made accessible by design. Chapter II applies to both 'personal' and 'non-personal data' without prejudice to "EU law on data protection and privacy" (Recital 7).[87]

Considering possible conflicts, Chapter X (Art. 35) of the Proposal clearly states that the *sui generis* right provided for in Article 7 of the Database Directive would not apply to databases containing data obtained from or generated by the use of a IoT product or a related service (Ch.II).[88] The wording of the exemption give rise to unclarities— most importantly about whether Article 35 is a limited exclusion confined to Article 4 and 5 of the Data Act proposal or a general statement about the ineligibility of IoT data to sui generis protection.[89]

Recital 63 of the proposed Data Act further states that data holders should exercise the *sui generis* right in a way that does not prevent public sector bodies from obtaining and sharing data in accordance with Chapter V of the Data Act. Chapter V mandates data exchanges from the private sector to the public bodies when a public interest-related need for such data emerges.[90]

4. IP protection pertinent to utilitarian (functional) elements

4.1 A general overview: computer programs, algorithms and ML models

ML-based decision systems are technical assemblages that include several tangible and intangible components, comprising of several sub-systems. They are not standalone black boxes, but massive networked entities pieced together and layered like *Lego* where multiple scales of processes are shaped

[86] The DGA states that where public sector bodies are holders of the sui generis right provided in Article 7(1) of the Database Directive, they should not exercise that right in a way which will prevent or restrict the re-use of data beyond the limits set by the proposed Regulation.
[87] Charlotte Ducuing, "Chapter II of the Data Act – Data control of users" in Charlotte Ducuing, Thomas Margoni and Luca Schirru (eds), *White Paper on the Data Act Proposal*, (2022) CiTiP Working Paper Series, 22-27, 23.
[88] The proposal also seems to disregard the judicially established distinction between the obtaining and creating of data (spin-off doctrine) where the latter is generally accepted to be outside of the scope of the sui generis right. Thomas Margoni, Thomas Gils and Eyup Kun, "Chapter X of the Data Act and the Sui Generis Database Right" in Charlotte Ducuing, and others (eds) *ibid*. 74-79.
[89] See above 3.3.1.
[90] Antoine Petel, "Chapter V of the Data Act - What is the European concept of "B2G data sharing" in the Data Act proposal?" in Charlotte Ducuing, and others (eds) *ibid*. 47-49.

by a number of economic, technical, social, factors.[91] Above, data and databases have been examined as the expressional ("literary") elements of these assemblages. Regarding their IP-eligible *utilitarian* elements, the below analysis is based on a three-partite taxonomy as *algorithms (algorithmic techniques)*, *ML models* and the embedding *computer code*.[92]

The establishment of transparency mechanisms, especially the construction and deployment of the software tools for audit and testing, may necessitate the reverse engineering or the implementation of the computer code or the essential parts of the ADM system. Below, the IP implications of these transparency measures are discussed within the framework of copyright law (accorded to the literary elements of computer programs (4.2)) and patent law where these systems or parts of them qualify as novel inventions (4.3).

4.2 Copyright protection of computer programs as creative expression

Software (computer code) is a pluralistic work which presents a dual nature having both expressive (literary) and utilitarian (functional) aspects.[93] This duality has given rise to an ongoing controversy as to the proper form and scope of IP protection of software.[94] In determining the possible application of IP protection to a computer code, it is crucial to differentiate between the utilitarian and literary (expressional) elements. While the former may be subject to patent protection, for the latter, EU law provides a specially tailored copyright regime, namely the Software Directive.

Unlike a picture, musical composition or a piece of scientific or literary writing, a software (computer program) is a functional tool in the sense that it is designed to carry out a certain task, e.g., calculating the sum of numbers from 1 to 100 (the *Gauss* method). This is the utilitarian aspect of the program which may be subject to patent protection independent of the way that the instructions are expressed to execute certain function. Apart from this functionality, the expressional dimension of a computer program also amounts to a "literary work" as rendered in a programming language (source code). The copyright protection afforded to the computer programs is limited to this expressional dimension which is distinct from the algorithmic techniques and the ideas underlying the computer code—just as a recipe could be expressed in numerous ways.

Under the EU Law, computer programs (both source code and object code) are subject to copyright protection as provided in the Software Directive. The Directive applies to the expression of a computer program in any form, including the preparatory design material, machine code, source code, and the object code (Art. 1(2)).[95] Unauthorised copying not only includes the literal(verbatim) copying of the code but also the appropriation of the essence of the programmer's way of expression to instruct the computer. However, this protection covers neither the functionality of the computer program nor the

[91] Bucher (n 2) 47.
[92] "The assemblage of [the] three elements (Data + Algorithm + Model) is proposed as a general diagram of machine learning" *See* Pasquinelli (n 8) 6.
[93] Ballardini (n 38) 27-62.
[94] Michael S. Keplinger. "Computer Intellectual Property Claims: Computer Software and Data Base Protection", *Washington University Law Quarterly*, (1977), 461.
[95] "[In the EU] the traditional reluctance to afford patent protection to computer programmes under the dictate of the European Patent Convention (EPC) has led to a system that tends to favour a broad scope of software copyright." See Ballardini (n 38) 29.

underlying algorithmic techniques.[96] Hence, the use of the parts of a software that do not amount to an appropriation of the creative expression of the programmer may not be prevented under the Software Directive.[97] For instance, the making of a computer programme which emulates another programme, without copying the other programme's code or graphics, is not an infringement. By the same token, while the particular implementation of a *communication protocol* in certain programming language is protected by copyright, the protection does not extend to the protocol as a sequence of instructions. Having said that, although distinguishing between literary and non-literary elements may be simple for artistic works, in case of software, expression is an integral part of the underlying idea or functionality and thus, not easily extricable. In case of non-literal copying of a copyrighted programme, ascertaining which part of the programme behaviour is expressive and which part is functional is a challenging task. As a consequence, copyright protection may occasionally be extended to software's utilitarian elements, protecting functional ideas in a 'patent-like' manner. [98]

Turning to the IP protection of *algorithms* and *ML models*, an initial conceptual and terminological clarification is necessary as these concepts are a constant source of confusion. In its current use, "algorithm" is a nebulous term with numerous subjective definitions the meaning of which frequently depends on the user and the context of the use. What we see is an imbroglio of perspectives and vocabularies, where concepts are easily misinterpreted, conflated, or used imprecisely.[99] In computer science, an algorithm is defined as a set of precise rules for solving a problem in a finite number of steps. These rules may contain logical operations, repetition, procession to another rule or temporary performance of another set of rules.[100] For the purposes of this paper, what we refer as "algorithm" is not this abstract outline of rules or instructions (based on common methods such as *Naive Bayes classifier, Linear Regression* or *K-nearest neighbours*) but its specified form as tailored for a given task within an ADM framework. The tailoring of the algorithm involves, for instance, specifying the number of trees and splits in a random forest algorithm or the depth of the hidden layers in a neural network.[101] An automated decision-making system may employ numerous algorithms, easily totalling to a number of hundreds or more.[102] Association of the output of one algorithm with the input of another is the standard method for composing algorithms into larger algorithms and eventually into ML models.[103] Once the best configuration to properly classify the training data is found, the resulting amalgamation

[96] Josef Drexl and others, "Data Ownership and Access to Data Position Statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the Current European Debate" (2016) Max Plank Institute for Innovation & Competition Research Paper 10, para. 14.
[97] In *SAS Institute Inc v World Programming Ltd* CJEU ruled that the functionalities of a computer program are not eligible, as such, to copyright protection and it will be for the national court to examine whether, in reproducing these functionalities, the author of the program has reproduced a substantial part of the elements of the first program. See C- 406/10 *SAS Institute Inc v World Programming Ltd* EU:C:2012:259.
[98] Guido Noto La Diega, "Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information" *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* (2018) 9(3), para 37. For a comprehensive treatment of this topic, especially in relation to the US Copyright Law, *see* Pamela Samuelson, "Functionality and Expression in Computer Programs: Refining the Tests for Software Copyright Infringement" *Berkeley Technology Law Journal* (2017), 31 (3):1215-1300.
[99] Liane Colonna, 'A Taxonomy and Classification of Data Mining' *SMU Science and Technology Law Review* (2013) 16, 309, 314.
[100] Kenneth Oksanen Perttu Virtanen, Eljas Soisalon-Soininen, Jukka Kemppinen, "Arguments in Considering the Similarity of Algorithms in Patenting" (2011) *SCRIPTed* 8(2):138-153, 139.
[101] Bucher (n 2) 20, 26.
[102] Thomas H. Davenport, *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work* (Cambridge, MA: MIT Press, 2018), 12.
[103] Oksanen and others (n 100) 139.

of algorithmic techniques (as specified and adjusted) comprise the ML model. As such, leaving aside their possible expression or representation in a tangible form, both the ML models and the algorithms as abstract formulas cannot conceivably amount to a solid expression within the sense of copyright law.[104] Nor could they easily satisfy the originality threshold since the functional constraints and applicable standards may severely limit the possibility of author's creative imprint.[105]

When copyright was considered as a measure of protection for software in the EU, it was not clear what role patents, trade secrecy, contracts or TPMs would play in the field of information technologies. In the end, the reluctance of the EU to afford patent protection to computer programmes under the dictate of the European Patent Convention (EPC) has led to a system that tends to favour a broad scope of software copyright. Accordingly, this broad scope together with the scarce and inconsistent case law of the CJEU give rise to an uncertain regime regarding the extent of protection on the computer programs.[106]

Exceptions and limitations

One of the most significant exceptions to copyright protection on computer programs is the case of reverse engineering. Computer science and software industry have established techniques to obtain some approximation of the corresponding source code through the analysis (reverse engineering) of the object code. As such, reverse engineering is a methodology which is also essential for the audit and testing of ML systems, and it usually requires the making of a reproduction or derivative of the copyrighted elements during the course of the process.[107]

The Software Directive permits reverse engineering (*decompilation)* in Articles 5 and 6 for limited purposes.[108] Article 5 provides that the *person having a right to use a copy of a computer program* is permitted to observe, study or test the functioning of the program in order to determine the underlying ideas and principles. Based on the idea/expression dichotomy that withhold copyright protection to abstract information, the object of the provision is to enable a lawful user to test, debug and modify the program as the normal use necessitates. For our current analysis, reverse engineering for the purposes of audit of the system or to scrutinize a specific decision may not be straightforwardly regarded as related to normal use of the program, i.e., loading, displaying, running, transmitting and storing.

Article 6 deals with *decompilation* for the purposes of developing interoperable products and services provided that the necessary information has not previously been made available. This is limited to the parts of the program necessary for interoperability and may only be carried out by licensees or by those having a right to use a copy of the program. Reverse engineering (*decompilation*) to achieve the interoperability of an independent software, e.g., for scrutinizing automated decisions, could benefit from this exception.[109] However, an important downside of the provision in terms of contestation is

[104] Pamela Samuelson, 'Why Copyright Law Excludes Systems and Processes From Its Scope of Protection' (2007) *Texas Law Review,* 85(7): 1921-1977.
[105] For different approaches to copyrightability threshold under English (skill and labour) and German (individuality' or 'creativity') copyright laws, see Ballardini (n 38).
[106] *ibid* 23.
[107] Jonathan Band, "The Global API Copyright Conflict" *Harvard Journal Of Law & Technology,* Special Issue Spring (2018) 31:615-636.
[108] The Software Directive omits the term "reverse engineering" both in Article 5 and 6.
[109] This will also include any data required to enable interoperability between programs. See Support Centre for Data Sharing (SCDS) "B2 – Analytical report on EU law applicable to sharing of non-personal data" DG

that one who is subject to an automated decision does not easily fall within the definition of a *person having a right to use a copy of the program*. ML systems are not computer programs which could have a "lawful user" within the sense of running an application on a device. Having said that, under a broad interpretation of the provision, the users of online services such as Google search engine could be regarded as lawful users under an implied license.

The Directive prohibits contractual clauses contrary to Article 6 or to the exceptions provided in Article 5(2) and (3). Unlike *sui generis* database protection, Article 7 of the Software Directive provides an exception for the prohibition on the circumvention of TPMs (Article 6 of the InfoSoc Directive) for reverse-engineering purposes.[110]

4.3 Patent protection

As copyright is not sufficient to protect the idea or the functionality underlying the ML model, the utilitarian (functional) elements of ML systems are frequently the subject of patent claims, allegedly qualifying as novel inventions.[111] Due to the above explained dual nature of computer code, the application of patent law in this domain still remains a controversial topic in many respects. Nevertheless, patents for computer-related inventions exist both in the US and in Europe, albeit in varying forms and degrees.

The most controversial issue regarding the patent protection of software centres around the question of patentable subject-matter. European Patent Convention (EPC)[112], as the main legal framework of patent protection in the EU and other signatory countries, accepts the patentability of computer-implemented inventions but does not recognise abstract computer code as patentable subject-matter. This is formulated in the wording of Article 52(2)(c) as excluding computer programs from protection "as such". It is understood from the case law of the European Patent Organisation (EPO)[113] that the interpretation of the "as such" exclusion presents difficulties.

The Convention classifies computer code, algorithms, ML models under "mathematical methods" as an excluded category of subject-matter.[114] Under the Convention, in order to be regarded as inventive, the computer code must provide a specific "technical solution to a technical problem". Inventions relating to ML and artificial intelligence (AI) are examined in the same way as inventions embodying mathematical methods. Hence, the aspects of the computer code which do not make any concrete contribution to the technical character of the invention are disregarded in the assessment of patentability.

CONNECT, SMART 2018/1009 24 January 2020 V2.0., 38. However, the reverse engineering exceptions of the software Directive do not extend to acts for establishing *data interoperability.* On this matter, in connection with the copyright protection of *application programming interfaces* (APIs), see Drexl, (n 46) 87.

[110] See above 3.2.2

[111] In a ML system, there could also be other independent patent claims directed to the methods for training, structuring or transforming data. Sam Jones, "Patentability of AI and machine learning at the EPO", (*Kluwer Patent Blog,* 21 December 2018) <http://patentblog.kluweriplaw.com/2018/12/21/patentability-of-ai-and-machine-learning-at-the-epo/?print=print>

[112] European Patent Office, *European Patent Convention* (17thedition, November 2020) <https://www.epo.org/law-practice/legal-texts/epc.html>

[113] EPO is the intergovernmental body administering the EPC.

[114] Article 52(2) of the EPC.

In recognition of the increasing number of ML-based applications, in its *Guidelines for Examination*, EPO provides some clarity on the matter.[115] According to the Guidelines, artificial intelligence and machine learning are based on computational models and algorithms for classification, clustering, regression and dimensionality reduction. Such computational models and algorithms such as neural networks, genetic algorithms, support vector machines, k-means, kernel regression and discriminant analysis are per se of an abstract mathematical nature, irrespective of whether they can be "trained" based on training data.

Accordingly, algorithms or the computer code which do not specify the use of any technical means are excluded from patent protection. [116] For instance, a classification algorithm without any indication of a specific technical use will not be regarded as having a technical purpose.[117] The ML model (including the internal structure and the specifications of the system), the algorithms and the implementing computer code may be regarded as inventive where the subject-matter is tied to a technical effect.[118] Patents on ML- based inventions are permissible so long as they are described in relation to operation or control of an apparatus or a process. So far, EPO has accepted inventions for detecting persons in a digital image, estimating the quality of a transmitted digital audio signal, separation of sources in speech signals, speech recognition, digital video enhancement and medical diagnosis as having technical character.[119] The same rules also apply to patent claims directed to methods of simulation, design or modelling which fall under the category of mathematical methods or of methods for performing mental acts. On the other hand, methods based on *natural language processing* (NLP), e.g., for detecting "junk" email, have generally been regarded to lack the necessary technical inventive step. This is because in a textual analysis, the relationship between the input and output is deemed to be not of technical nature but rather related to the *"*abstract linguistic information content*"*.[120] Accordingly, ML-based recommendation or rating systems are also found to be devoid of inventive step. The EPO asserts that, from a "technical" point of view, it is irrelevant what songs, videos, restaurants, hotels or etc. are recommended to a user.[121] However, it should be noted that the decisions of the EPO have

[115] EPO *Examination Guidelines* 2022, Part G, Chapter II, 3.3.1 (Artificial intelligence and machine learning) <https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_ii_3_3_1.htm>

[116] "[...] patent protection is reserved for inventions involving a "technical teaching", i.e. an instruction addressed to a skilled person as to how to solve a particular technical problem using particular technical means." EPO *Examination Guidelines* 2022*,* Part G, Chapter II (Inventions) 2.

[117] See EPO Boards of Appeal, T1784/06 *Classification method/COMPTEL* of 21 September 2012, EP:BA:2012:T178406.20120921.

[118] "The claim is to be functionally limited to the technical purpose, either explicitly or implicitly. This can be achieved by establishing a sufficient link between the technical purpose and the mathematical method steps, for example, by specifying how the input and the output of the sequence of mathematical steps relate to the technical purpose so that the mathematical method is causally linked to a technical effect." EPO *Examination Guidelines* 2022, Part G, Chapter II, 3.3 (Mathematical methods).

[119] Guidelines illustrate the use of a neural network in a heart monitoring apparatus for the purpose of identifying irregular heartbeats as having a technical contribution. Part G, Chapter II, 3.3.1 (Artificial intelligence and machine learning).

[120] Philip Cupitt, "Patenting Artificial Intelligence at the European Patent Office" (*Marks&Clerk,* 11 April 2019) <https://www.marks-clerk.com/Home/Knowledge-News/Articles/Patenting-Artificial-Intelligence-at-the-European.aspx#.YDwfAS2cY1I>.

[121] The EPO Board of Appeal has confirmed this position (T 0697/17 (SQL extensions/Microsoft Technology Licensing, 17.10.2019) by stating that improvements to obtain semantically "better results" could not be regarded as a technical effect because the concept of "better search" was subjective in the context of retrieval based on semantic similarity. <https://www.epo.org/law-practice/case-lawappeals/recent/t170697eu1.html>

not been fully consistent on this matter and throughout the years, there have been various cases granting patent rights to core AI applications.[122]

Regarding computer programs, the Convention also excludes them from patentability under Art. 52(2)(c) and (3). Similar to AI and ML applications, computer programs are patentable where they produce a "further technical effect" when run on a computer. According to the *Examination Guidelines,* further technical effect is understood to be going beyond the normal physical interactions between the program (software) and the computer (hardware). The control of a technical process or the internal functioning of the computer itself or its interfaces are examples of technical effects which confer technical character to a computer program.[123] A computer-implemented data structure or data format may also have technical character as a whole and thus found to be eligible to patent protection. Data formats and structures contribute to the technical character where they have an intended technical use and they cause a technical effect when deployed according to this intended technical use.[124]

There are some further points worth considering when evaluating the patent-based impediments to transparency. *First*, the application of conventional ML or AI techniques does not suffice to pass the inventiveness test even if the problem that is being solved is novel and technical. *Second*, patent claims that are purely result-focused and functional fail to specify a technical solution but rather cover any solution to a generic problem.[125] *Third*, an ML model may be too tightly specific to a certain training data and thus, incapable of being sufficiently generalized to yield useful results with other training data. In such cases, the inventive step cannot be explained in concrete technical terms without reference to the training data, and this makes the identification of the kernel of the invention very difficult or impossible.[126]

Restrictions to patent protection

Unlike copyright law, patent system provides a much lesser catalogue of general exceptions and limitations.[127] There are few options in patent law doctrine and in the international treaties that may support the unauthorized utilisation of patented ML inventions for the purpose contesting automated decisions.

"Experimental use" of the patented invention includes activities to test a hypothesis or to assess whether an invention works as presented. For instance, utilisation of a patented invention in order to

[122] For instance, European Patent No. 0554083B1 dated 1999 relates to a "neural network*"* that learns a probability density for linking input data to output data without reciting any details of how the system is implemented in hardware. It is argued that, under the New Guidelines, the application would have been rejected on the ground that it is purely mathematical and devoid of technical character. Cupitt (n 120).
[123] *Examination Guidelines* 2022, Part G, Chapter II, 3.6 (Programs for computers).
[124] *Examination Guidelines* 2022, Part G, Chapter II, 3.6.3 (Data retrieval, formats and structures).
[125] Brian Higgins, "The Role of Explainable Artificial Intelligence in Patent Law" *News and Analysis of AI Tech Legal Issues*, 16 December 2018) <http://aitechnologylaw.com/2018/12/explainable-ai-crucial-in-this-area-of-law/>.
[126] Jones (n 111).
[127] The EPC does not provide any exceptions for the patentee rights because the Treaty is concerned only with the regulation of the grant of patent rights. At the EU level, Article 27(b) of the failed proposal for an Agreement relating to Community Patents (89/695/EEC) provided exceptions (e.g, private, non-commercial or experimental use) that are partially adopted by some member states. Sean M. O'Connor, "Enabling Research or Unfair Competition? De Jure and De Facto Research Use Exceptions in Major Technology Countries" in Toshiko Takenaka (ed), *Patent Law & Theory: A Handbook Of Contemporary Research, (*Cheltenham, UK; Northampton, MA: Edward Elgar Publishing, 2009), 519-567, 530.

submit information for regulatory approval (e.g., to produce and market the same pharmaceutical compound once the patent expires) may be considered as a non-infringing use.[128] In what follows, arguably, the individual who is subject to an automated decision could have an interest in demonstrating that the algorithm-related invention does *not* work as specified in the patent application.[129] However, it is unlikely that this exception could cover the implementation of a patented computer code for the purposes of developing software tools enabling or facilitating the contestation of ADM systems.

As an international framework for IP rights, TRIPs Agreement in Article 30 provides that limitations to patent rights are permissible on the condition that such restrictions do not unreasonably conflict with the normal exploitation of the patent and unreasonably prejudice the legitimate interests of the patent owner, while also taking into account the legitimate interests of third parties. Echoing the "three-step test" in Article 13, these conditions under TRIPs are cumulative—each being a separate and independent requirement. Failure to comply with any of the three conditions results in the exception being disallowed. TRIPs provides a further exception in Article 31(b) which permits uses by the governmental authorities or third parties authorized by law provided that the scope and duration of use is limited to the purpose. The provision reflects a lenient attitude for public uses that are of non-commercial nature.

Overall, in terms of transparency, patent protection is a relatively minor obstacle due to the disclosure requirement, and also for the reason that patent protection is only applicable in case of a novel inventive step. Having said that, in cases where a patented invention is to be used for the development of software tools aiming to scrutinise ADM system, this will require a case-by-case analysis to determine whether any specific use of a patented ML technology jeopardises the legitimate interests or expectations of the patent holder. Like other IP types, patent protection too does not aim at a total control of information but simply grants a priority right in relation to the commercial exploitation of certain technical knowledge.

5. Trade secret protection

5.1 Trade secret: An alien species in IP law

The above-described fragmented landscape of IP rights with an inconsistent set of exceptions and limitations give rise to several ambiguities as to the IP protection of ML elements. As explained above, patents are subject to strict disclosure requirements and eligibility conditions and they just last for 20 years including the time spent to transform the invention to a marketable product. Speaking of copyright, although the copyright regime grants protection for a longer time span, it protects only the form of expression and does not extend to the functionality of the system. More importantly, many of the ML elements (e.g., ML models, algorithms, system specifications, statistical values, metrics used to

[128] Similar "Bolar exemptions" now exist in many countries in different forms. Some (e.g., United States) are solely confined to pharmaceuticals, while others are broader. The Canadian, Egyptian, Indian, Israeli, and Japanese exceptions, for example, are not industry specific. Lionel Bently and others, "Study on Exclusions from Patentability and Exceptions and Limitations to Patentees' Rights" (WIPO Standing Committee on the Law of Patents, SCP/15/3 Annex I, 2010), 33 https://www.wipo.int/edocs/mdocs/scp/en/scp_15/scp_15_3-annex1.pdf >

[129] La Diega (n 98) para. 41.

calculate probabilities, reference groups or the IT infrastructure) are not IP-eligible. Considering these mismatches and shortcomings in the IP protection, this Part explains to what extent trade secret regime could offer an escape route against transparency demands.

As opposed to distinct segments of IP protection analysed in the previous parts of this paper, no specific categories exist for defining the subject-matter eligible for trade secret protection**.** Any information of business value can qualify as a trade secret. The rationale of the trade secret protection is principally to promote commercial ethics and encourage innovation. It is also argued that a legal entitlement to trade secret protection avoids excessive efforts to prevent access to vital information as it cuts down the costs that would otherwise be incurred by zealous pursuit of "real secrecy".[130] Trade secret law reinforces both the physical and the contractual restrictions that businesses deploy to safeguard information.[131] Considering the constricted and uncertain nature of software and database protection, this broad scope of protectable information makes trade secret the preferred form of "appropriability mechanism" for ML-based systems—often combined with technological protection measures and contractual arrangements.[132]

As an odd member of the IP family, trade secret law is a unique type of protection which directly aims at the concealment of information.[133] Trade secret protection allows businesses to control information which is not eligible for patent protection (e.g., for lacking technical effect or inventive step) or where businesses are unwilling to disclose the inner workings of their system through a patent application.[134] However, trade secret regime is not necessarily a substitute for patent protection as both legal regimes may be put to use in a complementary fashion.[135] For instance, in addition to the patent protection over the PageRank's original algorithm, trade secret law protects all subsequent adjustments and specifications made by *Google*. The continual tweaking of the algorithm is rather kept in the dark to defeat those who intend to gain salience in search results.[136]

As a catch all framework, trade secret law creates economic incentives by prohibiting the appropriation or the use of commercially valuable information via unlawful means.[137] As explained in

[130] "For example, rather than triple-locking every vault or biometrically assessing the credentials of all who seek access, a trade secret owner can bind employees, customers, and others not to misappropriate or disclose valuable processes and product." Pasquale, (n 25) 244.
[131] Mark A. Lemley, "The Surprising Virtues of Treating Trade Secrets As IP Rights" (2008) *Stanford Law Review* 61(2): 311-351, 313.
[132] Wendy Seltzer, "Software Patents and/or Software Development" (2013) *Brooklyn Law Review* 78(3): 929-987; Andrew Beckerman-Rodau, "The Choice Between Patent Protection and Trade Secret Protection: A Legal and Business Decision" (2002) 84 *Journal of the Patent & Trademark Office Society*, 371.
[133] Because trade-secret law aims for suppression of information, many do not regard trade secret as a genuine IP right for being contrary to the basic rationale of dissemination of knowledge/information.
[134] Robert G. Bone, "The (Still) Shaky Foundations of Trade Secret Law" (2014) 92 *Texas Law Review,* 1803. On what amounts to sufficient disclosure in the context of software patents, *see* Tomkowicz (n 17) 33.
[135] Accordingly, it is also argued that because of the "fairly weak disclosure rules, a patent applicant can often secure a patent without disclosing all of the technologically and commercially important details of an invention. This withheld knowledge can in turn be protected under trade secret law." Brenda Simon and Ted Sichelman, 'Data-Generating Patents' (2017) 111 *Northwestern University Law Review,* 377-437, 384-389.
[136] "The legitimate reasons for search engines' general emphasis on keeping ranking algorithms confidential throw some light on the divergent rationales for adopting patent or trade secrecy protection for any given instance of intellectual property." Frank Pasquale, "The troubling consequences of trade secret protection of search engine rankings" in Rochelle C. Dreyfuss and Katherine J. Strandburg (eds) *The Law and Theory of Trade Secrecy* (Cheltenham, UK ; Northampton, MA: Edward Elgar, 2011), 381-405, 386.
[137] Jeanne C. Fromer, "Machines as the New Oompa-Loompas: Trade Secrecy, the Cloud, Machine Learning, and Automation' (2019) 94 *New York University Law Review*,706, 712.

this *paper*, what makes trade secret a flexible thus an attractive solution is that, even where IP rights are not applicable, certain information may still be protected against illicit misappropriation under TS Directive. This "blanket" cover raises concern in that excessive reliance on TS protection could give rise to disincentives for potential investors as it creates uncertainties with regard to the validity and enforceability of data transactions. As such, trade secret law may be seen as the most important IP impediment for transparency requirements in connection with ADM.[138] This is increasingly the case where public agencies outsource data collection and analysis tasks to private entities which rely on trade secrets as a part of their business strategy.[139]

5.2 General legal framework

The TRIPs Agreement, as the most comprehensive international framework for IP protection, provides that holders of undisclosed information are entitled to prevent "the information lawfully within their control from being disclosed to, acquired by, or used by others without their consent in a manner contrary to honest commercial practices."[140] With an explicit reference in Article 39, TRIPS Agreement treats trades secrets under the umbrella of "unfair competition".[141] The Article lays out the minimum level of protection and provides three conditions for an information to qualify as trade secret: i) secrecy; ii) commercial value; and iii) reasonable steps to keep the information secret.[142] As to the implementation of the Article by the signatory states, there is no consensus about the legal nature of trade secret protection, and countries have adopted different approaches. While some jurisdictions perceive trade secrets as a special case of unfair competition or tort law (e.g., breach of confidence), some rely entirely on contract, excluding any property-based approach due to concerns that a conception of property could result in excessive protection.[143]

At the EU level, trade secret protection is harmonised by the Trade Secrets Directive.[144] The Directive provides a minimal standard of protection and refrains from obliging member states to recognize property-based rights.[145] It is a framework allowing Member States to maintain their preferred type of

[138] "[...] secrecy has also compromised inquiries into the validity of factual determinations made by voting machines and intoxication-detection instruments. Both judicial decisions and secondary literature have investigated the degree of secrecy needed in these fields in order to balance the proprietary rights of software owners and the right of the public to know exactly how given actions have been interpreted by machines." Frank Pasquale, (n 136), 382.
[139] *ibid*. 398.
[140] The term "trade secret" is not used in the TRIPS. Also *see* Convention of the Union of Paris, Paris Convention For The Protection Of Industrial Property (As Amended On September 28, 1979) <https://www.wipo.int/treaties/en/text.jsp?file_id=288514>
[141] Gustavo Ghidini and Valeria Falce, "Trade secrets as intellectual property rights: a disgraceful upgrading – Notes on an Italian 'reform'" in Dreyfuss and Strandburg (eds.) (n 136) 140-151, 141.
[142] Hanns Ullrich Reto M. Hilty, Matthias Lamping, and Josef Drexl (eds), *TRIPS plus 20: From Trade Rules to Market Principles* (Berlin, Heidelberg: Springer, 2016); Mira Burri and Ingo Meitinger, "The Protection of Undisclosed Information: Commentary of Article 39 TRIPS" in Thomas Cottier and Pierre Véron (eds), *Concise International and European IP Law: TRIPS, Paris Convention, European Enforcement and Transfer of Technology* (The Hague: Kluwer Law International, 2014).
[143] Banterle, 'The Interface (n 50) 416.
[144] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure.
[145] On whether trade secrets are a form of property under the *European Convention of Human Rights* and the *EU Charter of Fundamental Right*s, Aplin argues that the Directive does not provide a robust property approach and instead adopts an unfair competition model. Tanya Aplin, "Right to Property and Trade Secrets" in Christophe

protection as long as undisclosed know-how and business information are safeguarded against misappropriation.[146] The formulation used in the Directive follows the wording of the TRIPs with a broad approach to the notion of trade secret holder. According to Article 4(2a) of the Directive, unlawful acquisition of a trade secret means "unauthorised access to, appropriation of, or copying of any documents, objects, materials, substances or electronic files, lawfully under the control of the trade secret holder, containing the trade secret or from which the trade secret can be deduced". Acquisition of trade secrets through other conducts which are considered contrary to honest commercial practices is also prohibited under Article 4(2b).

Below parts explore which of the functional and expressional elements of ADM systems fit within the Directive's general understanding of information and what may be the exceptions or limitations to trade secret protection for the sake of algorithmic transparency.

5.3 Data and database as trade secret

Individual data items contained in a dataset could satisfy the secrecy requirement under Article 2(1a) of the Trade Secrets Directive in the sense of being (i) not generally known or (ii) not readily accessible. While generally known factual information such as one's age, gender etc. may be regarded not to satisfy the secrecy requirement, information such as the exact location of a pothole on the city roads (as being known by many of the citizens) poses a more difficult question—not lending itself easily to a straightforward answer.[147] Leaving this aside, data generated, for instance, by the heat sensors in a machine would qualify as a trade secret for containing valuable information about the manufacturing process.[148] Subject to the restrictions of the GDPR, information derived from data (e.g., one's eating habits, health situation) could also enjoy trade secret protection.

Other than individual data items, databases also enjoy trade secret protection notwithstanding whether they are eligible to *sui generis* right or copyright protection.[149] The source of the data, whether it is obtained from individuals, measured by sensors, generated in a machine-to-machine process or captured through tracking technologies does not have a bearing on the evaluation of the secrecy requirement in relation to a dataset.[150] According to Article 2(1) of the Directive, in terms of secrecy requirement, a database is treated as a unit in its entirety. The Article provides that the information

---

Geiger (ed), *Research Handbook on Human Rights and Intellectual Property* (Cheltenham, UK: Edward Elgar, 2015), 421-437.

[146] Gintarė Surblytė, 'Enhancing TRIPS: Trade Secrets and Reverse Engineering' in Ullrich, and others (eds) (n 142), 725-760, 726. Recital 2 of the Directive describes trade secret protection as a 'complement' or an 'alternative' to IP rights. In line with this, enforcement of trade secrets is not subject to the Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights, which harmonizes IP Enforcement throughout the EU. See Drexl, (n 46) 91.

[147] In their position statement, scholars from Max Planck Institute have expressed concern that it was decisive whether the factual exclusivity of data fell under the scope of the Trade Secret Directive. Drexl and others, (n 96) para. 21.

[148] Herbert Zech, "A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data", (2016) *Journal of Intellectual Property Law & Practice* 11(6): 460–470.

[149] For example, Google treats the complete list of titles in its *Google Books* project as "commercial intelligence". Since the company does not want to disclose which of the books it has so far scanned, sui generis database right or copyright, which do not allow for secrecy but rather prevent others from extracting or re-utilising the content in certain ways, do not offer a satisfactory solution.

[150] On trade secret protection of the data gathered via the Internet of Things, *see* Cristiana Sappa, "What Does Trade Secrecy Have to Do with the Interconnection-based paradigm of the Internet of Things?" (2018) *European Intellectual Property Review*, 40(8):518-523.

must be secret "in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known". For instance, the aggregated customer data as a whole may well qualify as a trade secret, even if each individual customer information can be found in open sources. Independent of the secrecy of the individual data items, secrecy with respect to a database also includes how individual pieces of information relate to each other.[151] Accordingly, a personal profile comprising of many selected, possibly trivial, factual data may possess the necessary quality of confidentiality.[152] This also applies to datasets enabling the deduction of an information protected by trade secret.[153] As Drexl draws attention, data can have very different functions based on the specific interest of the person who is seeking access.[154]

Datasets are treated confidential both due to the information they contain (e.g., the secret formula of Coca-Cola) and also for their function in a certain process. In the latter case, e.g., a properly labelled training dataset, the claimed protection does not aim to maintain the confidentiality of semantic information but rather to deprive the competitors from a useful asset or tool. In ML context, the utility of a database lies in the statistical correlations contained in larger sets of aggregated data. Accordingly, training data should rather be considered as a kind of resource, distinct from the concrete semantic information it embodies.[155]

Regarding the commercial value requirement in Article 2(1b), the Directive does not set a threshold but rather deems commercial value implicit based on the investment made for obtaining or generating data or the mere effort to keep it secret. Even if the publicly available data might not possess commercial value, they may nonetheless provide a competitive advantage when compiled into a database. As Recital 14 of the Directive states that value could be actual or potential, unstructured data could also be of commercial value.[156] In general, the existence of a market is likely to be *prima facie* evidence for the "worthiness" of the data. Yet, as Drexl put it: "[…] while data may nowadays have great commercial value, it is quite questionable whether it will always be possible to establish a causal link between the secrecy of the information and its commercial value."[157]

Trade secret law grants legal protection to *de facto* secrecy. The requirement of reasonable steps to keep information secret could be achieved both by technical and organizational measures such as designating restricted areas in the company premises or introducing access restrictions. Businesses also heavily make use of contractual clauses mandating confidentiality or precluding reverse-engineering.[158] The adequacy of the measures is relative and will be determined in consideration of the trade secret holder's economic size, sectoral conditions, prior experience with trade secrets and the organisational policies. In this regard, the reference made to reasonableness and specific circumstances in Article

[151] Drexl, (n 46), 93.
[152] Drexl and others, (n 96) para.25. Zech also points "with Big Data, trivial information can have economic value when there is enough trivial information put together and analysed. Zech, (n 148), 460.
[153] For a discussion whether the new information derived from the analysis of a dataset protected as trade secret could be regarded as "infringing good" (Art. 2(4) and Article 4(5)), see Drexl, (n 46), 98-99.
[154] Josef Drexl, 'Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy' in Alberto De Franceschi and Reiner Schulze (eds), *Digital Revolution – New Challenges for Law* (C.H.Beck and Nomos 2016).
[155] Drexl, (n 47), 281, para 127.
[156] Nuno Sousa e Silva, "What Exactly Is a Trade Secret under the Proposed Directive?' *Journal of Intellectual Property Law & Practice* 9(11) (1 November 2014): 923–932, 924.
[157] Drexl, (n 47), 269, para 54.
[158] Mariateresa Maggiolino, "EU Trade Secrets Law and Algorithmic Transparency" (2019) Bocconi Legal Studies Research Paper No. 3363178, 9.

2(1c) pinpoints a test of proportionality.[159] In ML applications, particularly in case of analysis of real-time data, there could be special difficulties in the enforcement of the secrecy measures arising out of intricate supply chains and numerous participants which make contractual arrangements impractical or too costly.[160] In such cases, system owners and operators also resort to technical solutions such as digital encryption to keep their data secret.[161]

5.4 Trade secret protection of utilitarian (functional) elements

Considering neither the copyright nor the patent protection shields against transparency demands, it is common practice that functional (operational) elements of the ML systems are also technically and/or contractually guarded, and thus treated as trade secret. Trade secret protection, as having the widest scope to embrace virtually any kind of information, could provide the desired control over the ML elements where patent or copyright protection provides no satisfactory solution. Several aspects of machine learning technology (algorithms, ML models and the computer code) fit within trade secret law's general understanding of information.[162] Both the *literary* and *utilitarian* dimension of algorithms could enjoy trade secret protection. What may be protected is not only the functionality of the algorithm but also the exact sequence and the precise form of the instructions and mathematical formulations on which it has been built upon. It is common practice that software vendors license their object code but keep the source code secret via specific contractual clauses, prohibiting reverse engineering or otherwise *decompilation* of the object code.

Unlike other IP types, trade secret protection covers abstract information such as a process, mechanism, model, or a profile. Under Article 2(1a), a ML model will be considered secret where it is not, as a body or in the precise configuration and assembly of its components, generally known.[163] Furthermore, many different types of information such as the source of the training data, data features and weights, data structures, metrics and other internal parameters could be subject to trade secret protection. This may even include the exact configuration of the hardware or the software packages used.

5.5 ML output and trade secrets

It has already been mentioned above that where ML output is accessible, it is possible to estimate the essential properties of the model by sending specifically designed queries.[164] For instance, inputs could be tailored to discover whether certain individuals or groups were included in the training data.[165]

---

[159] Robert G. Bone, "Trade Secrecy, Innovation and the Requirement of Reasonable Secrecy Precautions" in Dreyfuss and Strandburg (eds), (n 136) 46-77.
[160] It is generally accepted that the Database Directive is not equipped to address the ambiguity as to the legal status of data as an economic resource. See Commission, 'Staff Working Document—Executive Summary of the Evaluation of Directive 96/9/EC on the legal protection of databases' SWD (2018) 146 final.
[161] Dag Wiese Schartum, "Making privacy by design operative" (2016) *International Journal of Law and Information Technology* 24(2):151–175.
[162] Michael Mattioli, 'Disclosing Big Data' (2014) *Minnesota Law Review* 99(2): 535-583.
[163] Drexl and others, (n 96) para.25.
[164] Florian Tramer, Fan Zhang, Ari Juels, Michael K. Reiter and Thomas Ristenpart."Stealing Machine Learning Models via Prediction APIs" (25th USENIX Security Symposium, USENIX Association, 2016) 601–618. <https://floriantramer.com/docs/papers/sec16stealing.pdf>
[165] Micheal Veale Reuben Binns, and Lilian Edwards, "Algorithms that remember: model inversion attacks and data protection law" (2018) *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 213. < http://dx.doi.org/10.1098/rsta.2018.0083>

Combined with the training data, ML output also enables insights into the inferential logic and the decisional criteria employed by the system. Hence, the analysis of a large enough sample of results may be necessary to prove unfair, discriminatory or otherwise unlawful consequences at several stages of ADM (i.e., selection of the training data, training of the algorithms or the analysis carried out for a specific decision). Without access to output data, the reverse engineering of the ML-based systems could be severely restricted.[166] Where it prevails, trade secrecy makes it practically impossible to test whether the outcome of ML in the form of ratings, scores, prescriptions or forecasts are correct, accurate or reliable.[167]

Considering the broad scope of the protection provided in the Trade Secrets Directive, it is likely that the output of the ML analysis either in machine-readable data form or as semantic information will be covered by the EU legislation.[168] It is no surprise that developers and operators of ADM systems frequently resort to trade secret claims to prevent access to the output of their systems. Simon and Sichelman draw attention that trade secret protection over the data generated by a patented ML-based invention would significantly extend the IP monopoly. This enables the patentee to leverage the reams of data generated throughout the exploitation of the patented technology for further competitive advantage.[169] Since patent law does not contemplate to cover information generated by the invention, such excessive protection is criticised in that it could result in a negative impact on downstream innovation and increase deadweight losses.[170] When combined with possible patent rights relating to the ML model or algorithms, trade secret protection of the output data could broaden the effect of the patent beyond the scope intended by the legislature.

5.6 Relevant limitations under the Trade Secrets Directive

The reference to "honest commercial practices" in the Trade Secret Directive confirms that, in line with the TRIPS agreement, the Directive treats the infringement of trade secrets as a special form of tortious conduct also referred to as *unfair competition*. The Directive does not grant a property right in the information but rather establishes a liability regime. This makes trade secret a more modest protection than what would be enjoyed under the classic ownership theory.

The protection is not unconditional but subject to trade secret holder's strict preservation of the *de facto* secrecy. A trade secret will cease to exist even when the information is made public via illegitimate means without the consent of the right holder.[171] Accordingly, as a built-in limitation, the holder of a trade secret cannot prevent rival businesses, or in general third parties, from reaching out to figure out the undisclosed information through independent efforts. Despite its affordances in terms of

[166] Mattioli (n 162), 484.
[167] Pasquale, (n 25) 237.
[168] European Data Protection Supervisor, "Opinion of the European Data Protection Supervisor on the Proposal for a Directive of the European Parliament and of the Council on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) against Their Unlawful Acquisition, Use and Disclosure" (2014), 3 <https://edps.europa.eu/sites/edp/files/publication/14-03-12_trade_secrets_en.pdf>; Wachter and Mittelstadt (n 13).
[169] Brenda Simon and Ted Sichelman, "Data-Generating Patents" (2017) *Northwestern University Law Review,* 111: 377-437, 408, https://ssrn.com/abstract=2753547. For more on this and relevant German court cases, see Drexl, (n 46), 87-88.
[170] Hyunjong Ryan Jin, " Think Big! The Need For Patent Rights In The Era Of Big Data And Machine Learning", *NYU Journal of Intellectual Property and Entertainment Law*, 7(2): 78; Drexl, (n 47) para.61
[171] Maggiolino (n 158), 11. This is irrespective of any monetary compensation for damages that may arise from the unlawful appropriation or disclosure of the secreted information.

concealing information, the weak spot of trade secret protection is that the obtaining of the information covered by the trade secret by way of inspection or reverse engineering is permissible. Under Article 3(1b) of the Directive, the observation, study, disassembly or testing of a product or object that has been made available to the public or that is *lawfully in the possession of the acquirer* is not prohibited. As Recital 16 clarifies, this aims to prevent overprotection which impedes innovation and competition in the relevant market. Such limitation is also a necessary constraint to prevent conflict with the fundamental rights of freedom of speech and the freedom to conduct business.

Trade Secrets Directive requires that the person acquiring the secret through independent means should be free from any contractual bounds limiting such acquisition— meaning that reverse engineering is only allowed as long as there is no contractually or legally valid duty to the contrary. This allowance of contractual restrictions severely diminishes the effect of the limitation permitting independent discovery and reverse engineering as an implicit constraint of trade secret protection. Such contractual prohibitions are likely to bring about anticompetitive and counterproductive results.[172]

Allowing reverse engineering and independent discovery as legitimate paths toward learning a trade secret, the restriction has certain resemblance to the permitted uses of computer programs under the Software Directive— albeit with significant contrasts between the two regimes.[173] While Trade Secrets Directive permits contractual terms to prevent the reverse engineering of the trade secreted parts of their system, the Software Directive clearly states that contractual provisions prohibiting the observation, study or testing of the program shall be null and void. Accordingly, considering the *lex specialis* status of the Software Directive, it may be argued even if the reverse engineering of the computer program under the Software Directive reveals trade secrets, contractual limitations will not be applicable.[174] The Software Directive prevails, provided that the reverse engineering is carried for the purposes specified in Articles 5 and 6 of the Directive.

Although a trade secret could be acquired through reverse engineering or other techniques of independent discovery, overlapping patents may still preclude the use of the secret elements learnt through reverse engineering. It is argued that because of weak disclosure rules, an applicant can often secure a patent without disclosing all of the technologically and commercially important details of the invention.[175] For instance, in case of software-related inventions, disclosure of the source code is not a strict requirement. That is, a functional description of the computer program may be accepted as sufficient since the coding (writing) of the program (based on the description provided in the patent application) is regarded to be a relatively straightforward task for a skilled expert. This may allow the patent owner to treat the undisclosed source code as trade secret, blurring the boundaries between the two IP regimes.[176]

Other than independent discovery, Article 1, as defining a negative scope, draws the boundaries of the protection under the Trade Secrets Directive and accordingly, Article 5 provides certain exceptions. What is of significance for our analysis (algorithmic transparency) is the limitations relating to the pursuit of public and legitimate interests. Article 1(2b) states that the Directive shall

[172] David D. Friedman, William M Landes, and Richard A Posner, "Some Economics of Trade Secret Law" (1991) *Journal of Economic Perspectives* 5 (1): 61–7, 62.
[173] See above Part 4.2.1 and 3.3.2 for "lawful use".
[174] Noto La Diega (no 98) para. 35.
[175] Simon and Sichelman (n 169).
[176] Tomkowicz (n 17).

not affect "the application of Union or national rules requiring trade secret holders to disclose, for reasons of public interest, information, including trade secrets, to the public or to administrative or judicial authorities for the performance of the duties of those authorities."[177] Accordingly, courts and public authorities during the course of administrative or judicial proceedings (e.g., relating to the right contest under the GDPR) may require trade secret holders to disclose certain essential elements of their system to evaluate the dispute in hand. This parallels with the limitations in Article 5(d) of the Trade Secrets Directive for the purpose of protecting a legitimate interest recognized by Union or national law and Article 5(b) which requires that trade secret protection will not be applicable for revealing misconduct, wrongdoing or illegal activity, aiming to protect the general public interest.

6. Conclusion: An uncertain regime of discontent and the road ahead

As seen, other than access and disclosure in the conventional sense, the practical exercise of the right to contest may entail certain IP-infringing acts. The legal implications of these acts will depend on: i) the extent of the use (copying, transforming, implementing) necessitated by the preferred mode of implementation; ii) the economic impact of the actual use on the legitimate rights and interests of the right holders; iii) the applicability of exceptions and limitations and whether they could be overridden by contract.[178]

Based on the findings of the paper, speaking of non-commercial activities aiming for the exercise of a statutory right (e.g., contesting automated decisions under GDPR Article 22/3), an immunity could be carved out from the existing exceptions and limitations of the IP regime to exempt certain acts and data usage for the purposes of algorithmic transparency. Such interpretation addressing transparency requirements for the purpose of scrutinizing automated decisions would neither conflict with the three-step-test nor with the rationale of the IP protection.[179] However, under the above-explored regimes, there are a number of shortcomings for the interpretation of the EU *acquis* in this direction.

First, a significant number of the exceptions and limitations under the EU IP laws are non-mandatory, leaving discretion to member states whether to implement the restriction in their domestic laws. The mandatory restrictions scattered in various legislative instruments are also far from providing an efficient immunity as member states do not have a uniform approach. The incoherence and the lack of coordinated efforts among the member states have so far precluded these provisions from having effective application.

Second, a general overview of the exceptions and limitations for scientific research reveals discrepancies among different IP regimes, giving rise to a non-harmonized situation within the EU. While Infosoc Directive Article 5(3) requires that scientific research should be the "sole purpose" (excluding any non-

[177] Lionel Bently and Brad Sherman, *Intellectual Property Law* (4th ed, United Kingdom: Oxford University Press, 2014), 1181.
[178] As this paper is limited to a macro-view of the possible IP-related obstacles in the implementation of the right to contest, how these potential conflicts could concretise in various ADM contexts is beyond the scope of the study. Such inquiry will require insights into the intricate contractual and organisational relations and a *case-by-case* analysis of each actual use or interaction with the relevant IP-eligible item or element.
[179] It is also argued that, in the absence of a fair use doctrine, non-expressive uses of copyrighted material for ML purposes do not constitute a "reproduction" under the EU copyright law. Daniel Schönberger, "Deep Copyright: Up - And Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML)" in Jacques De Werra (ed.), *Droit d'auteur 4.0 / Copyright 4.0, (*Geneva / Zurich, Schulthess Editions Romandes, 2018), 145-173.

scientific research), the same exception in the Database Directive does not have this requirement. The mandatory TDM exception in Article 3 of the Digital Single Market Directive also omits the word "solely".

Third, in many cases, the exceptions and limitations granted by the IP legislation are contractually overridable, rendering them to a significant extent ineffective. The overridability of statutory exceptions and limitations through contractual arrangements has been an ongoing debate for each IP regime.[180] On this crucial issue, the CJEU generally accepts the prevalence of statutory restrictions over contractual clauses unless contract or license terms to the contrary are expressly allowed by the member state law.[181] As Synodinou draws attention, this will depend on the way that the exceptions are formulated in the national statutes law together with the dominant attitude of the member state law regarding the place of the author and the justification of copyright law. Moreover, legal traditions vary as to their recognition and practice of the principles of freedom of contract and the autonomy of the parties.[182]

Fourth, the condition of lawful use (expressed in several provisions in a confusingly varying terminology) severely restricts the application of the relevant exceptions for scrutiny and contestation purposes.[183] In that regard, the TDM exception of the DSM Directive has further complicated the matter by introducing novel variations, i.e., *lawful access* and *lawfully accessible*.[184] Combined with varying approaches to freedom of contract in national laws, the condition of lawful use/access renders the relevant exceptions and limitations impracticable and ineffective.[185]

Fifth, the Database Directive is subject to increasing criticism due to its failure in addressing the legal implications of the data-driven practices.[186] It is clear that the Directive was not drafted in contemplation of the current technological advances in ML, Industry 4.0 or the Internet of Things. Especially, the issue of distinguishing between *creating* and *obtaining* data, and consequently whether machine-generated databases shall be protected, cannot be easily resolved under the current regime.[187]

---

[180] Lucie M.C.R. Guibault, *Copyright Limitations and Contracts: An Analysis of the Contractual Overridability of Limitations on Copyright* (The Hague; Boston: Kluwer Law International, 2002); Lucie M.C.R. Guibault, "Why Cherry-Picking Never Leads to Harmonisation: The Case of the Limitations on Copyright under Directive 2001/29/EC" (2010) 1 *Journal of Intellectual Property, Information Technology and E-Commerce Law,* 55, para 1.

[181] C-457/11 *Verwertungsgesellschaft Wort (VG Wort) v Kyocera and Others* [2013] EU:C:2013:426. More on this series of judgments, *see* Synodinou, (n 66) 31, fn.9.

[182] *ibid*. Also *see* Borghi and Karapapa (n 60) fn.80.

[183] "The lack of a clear EU definition of the terms "lawful user", "lawful use" and lawful or legal access make an assessment of the possibility of invoking these copyright exceptions, where these terms are employed, and, generally of the lawfulness of the users' acts, dependent on a mosaic of possible interpretations." Synodinou, (n 66) 36.

[184] Triaille and others (n 40) 110. See above Part 3.4

[185] There are arguments in favour of a uniform and dynamic definition of the concept of "lawful use" to strengthen the position of users who can enjoy the exception as a reinforced legal prerogative akin to a "user right". See Synodinou, (n 66).

[186] Eropean Commission, "Commission Staff Working Document—Executive Summary of the Evaluation of Directive 96/9/EC on the legal protection of databases" SWD (2018) 146 final; Drexl, (n 46) 67-85; Drexl, (n 47) para.52; Wiebe (n 38).

[187] For the upcoming Data Act which exempts IoT data from sui generis protection under certain conditions, see above Part 3.5.

According to Drexl, the concept of database, as defined by the Directive, is far too static to adequately respond to the features of constantly changing datasets and real-time data services.[188]

Sixth, trade secret protection, as including any information which may be physically or contractually concealed, often serves as a "blanket cover" against transparency demands. Considering the EU Trade Secret Directive, it could be said that the Directive has not been specifically drafted with the emerging data economy in mind. The few possible exceptions provided do not seem to embrace the economic, political and social function of data. Hence, it is argued that the legal framework provided by the Directive is not sufficiently robust to avoid unwanted restrictions to the free flow of data. The application of the Directive in a digital context is likely to present difficult questions which remain to be answered by the courts.[189] Such view is also endorsed by the European Commission as the Communication (Intellectual Property Action Plan) intends to revise Trade Secret Directive — based on the concerns that excessive reliance on trade secrets could give rise to disincentives for potential investors as it creates uncertainties with regard to the validity and enforceability of data transactions.[190]

In sum, the data operations aiming to render ADM systems transparent for the purpose contestation do not easily fit in the exceptions and limitations of the EU IP regime. Leaving aside the transaction costs necessary to identify and negotiate several different types of IP rights, relying on exceptions and limitations which will cover several possible infringements for the purpose of algorithmic transparency requires a far-fetching interpretation of the existing provisions. Therefore, without intense judicial and administrative interpretative intervention, relying on different exceptions will be fraught with uncertainties and obstacles.

Considering a possible all-encompassing exception (or an improved version of the TDM exception)—which could render lawful both the disclosures and other technical and administrative means for the purposes of transparency in ADM—would not provide help either. Such a general exception with a broad scope could turn out to be not of much use since it will leave open the question of how to keep the restriction of private rights at a proportionate level and thus, how to strike a balance among conflicting interests.[191]

[188] Drexl, (n 47) para.49.
[189] Drexl, (n 46), 97, 106-108. Also, *see* Tanya Aplin, 'Trading Data in the Digital Economy: Trade Secrets Perspective' in Sebastian Lohsse and others (eds), in (n 57).
[190] European Commission, "Making the Most of the EU's Innovative Potential, An Intellectual Property Action Plan to Support the EU's Recovery and Resilience" (Communication) COM(2020) 760 Final (21 November 2020) <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020DC0760> . The Data Act proposal repeatedly refers to trade secrets in various provisions stating that appropriate measures shall be taken to preserve the confidentiality of the trade secrets. However, these references provide almost no guidance about how the provisions of the proposed Act relating to access or sharing of data will be applied to cases where the data question contains or constitutes a trade secret.
[191] As the analysis of the paper is restricted to the potential areas of conflict between the IP regime and the transparency requirements, the question of conflict between the fundamental rights to property and data protection has been left untouched. The conclusion of the thesis (Part.3.2) briefly returns to this question illustrating certain difficulties about the application of the proportionality test of the CJEU.

Chapter 6

Conclusions

Conclusions

1. A prelude to the conclusion

As the thesis comprises of a series of papers scattered over a period of more than four years since the start of the PhD project, there has been a constant flow of academic articles, books and stakeholder initiatives which somehow formulate ways to render these systems *fair, accountable* and *transparent*. Scholars from various disciplines have been increasingly engaged with the issue of algorithmic transparency as a key-instrument for enhancing accountability of ADM systems, providing explanations to the affected individuals and thus limiting the adverse effects of their obscure nature.[1] Considering this, the conclusion not only reflects on the papers, but also attempts to refer to significant academic, judicial and legislative developments that have emerged after the publication of the relevant paper until the specified closing date of the thesis (January 26, 2023).

During this period, AI and ADM, together with the initiatives aiming to increase the availability and the reuse of data in B2B and B2G contexts, have been the top agenda items of the EU institutions—resulting in numerous communications, recommendations, guidelines and more importantly legislative developments such as the AI Act[2] (proposal), DSA[3] and DMA[4] as well as DGA[5] and DA[6] (proposal). This dynamic environment has made the research question of the thesis a constantly moving target. Hence, in addition to a summary of the findings and future projections, this conclusion also constructs links with the legislative developments aiming for the responsible deployment of AI technologies and the lifting of the restrictions which impede access to data. Through these links, a further aim of this

[1] Emilee Rader, Kelley Cotter, and Janghee Cho, "Explanations as Mechanisms for Supporting Algorithmic Transparency" in Proceedings of the 2018 *CHI Conference on Human Factors in Computing Systems - CHI '18*, ACM Press, 2018, 1–13 ; Cary Coglianese, and David Lehr, "Transparency and Algorithmic Governance" (2019). Faculty Scholarship at Penn Law 2123.https://scholarship.law.upenn.edu/faculty_scholarship/2123 ; Felzmann, Heike, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. "Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns" *Big Data & Society* 6, no.1 January-June 2019; Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi, "Fairness and abstraction in sociotechnical systems" In Proceedings of the *Conference on Fairness, Accountability, and Transparency*, pages 59–68. ACM, 2019; Frank Bannisterand Regina Connolly. "Administration by algorithm: a risk management framework. Information Polity, 2020, 25(4):471-490 ; Rik Peeters, "The Agency of Algorithms: Understanding Human-Algorithm Interaction in Administrative Decision-Making" 2020, 25(4): 507–522; Karen Yeung and Martin Lodge (eds.), *Algorithmic Regulation,* Oxford: Oxford University Press 2019; Ida Koivisto "Transparency in the Digital Environment" *Critical Analysis of Law,* 2021, 8(1): 1-8; Robert Herian, *Data: New Trajectories in Law,* Milton Park, Abingdon, Oxon ; New York, NY: Routledge, 2021.

[2] Proposal for a Regulation of the European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final} (hereafter AI Regulation);

[3] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act).

[4] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

[5] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). *See* above Ch.5, sec.3.5

[6] Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM/2022/68 final. Ch.5, sec.3.5.

chapter is to provide a more concrete assessment of the transparency model and the practical requirements elaborated throughout the thesis.

Given that each paper, as an independent research item, contains a conclusion extending beyond the relevant sub-research question, the conclusion below also aims to contextualise the compiled papers within the wider questions of (1) *how could law as a system absorb and internalise (contextualise) the use of algorithms (computer guided processes) for decision-making purposes and (2) what are the points of entanglement between law (rights, duties, prohibitions) and data-driven processes?* To this end, the below parts evaluate the research output in two sections, handling transparency in ADM and IP implications separately.

2. On the concept of transparency

2.1 Summary of findings

As the starting point, the first paper is an inquiry on the question of "*why transparency*", exploring what harms ensue from ADM which bring about the debate of transparency. As a prelude, the thesis defines ADM as a regulatory technology and identifies normative, causal and moral impairments which undermine the principle of rule of law. By showing the impairments to one's capability to reason with automated processes, the first paper sets the scene for further analysis on what interpreting the "algorithm" could mean for the purpose of contesting automated decisions. It establishes the perspective that by sorting, classifying and predicting, ADM systems may be regarded as imposing or facilitating certain norms or regulatory orders. This theoretical stance allows for a conceptualisation of ADM and the surrounding transparency debate as a procedural, or we may say, as a *due process* problem. This procedural approach forms the theoretical backbone of the thesis and paves the way for further analysis about what transparency entails in ADM and how these requirements could be implemented.

Having conceptualised ADM as a regulatory process and thus set the scene to approach transparency as a procedural problem, the thesis takes the view that the notion of transparency in ADM not only refers to barriers to access to information, but also concern the issue of interpretability of such information where understanding how the output has been generated is more of a challenge.[7] In this perspective, transparency in ADM is not conceptualised by the outcomes it is intended to bring about, but rather by the specific requirements focussing on what types of information and mechanisms could enable effective contestation.[8] Accordingly, the second paper focusses on the *"how of the transparency*", laying out a *contestation-specific* typology of transparency challenges (informational asymmetries) in ML whether they stem from corporate or state secrecy, technical illiteracy or from the lack of interpretability.[9] Based on this taxonomy, the transparency model developed in the *second* paper is a reconstruction of ADM akin to a 'rule-based' process where certain input leads, to certain results—akin to the decisions in a legal system based on *facts*, *norms* and the ensuing *consequences*.

---

[7] Francesca Palmiotto, "The Black Box on Trial: The Impact of Algorithmic Opacity on Fair Trial Rights" in Martin Ebers and Marta Cantero Gamito, (eds), *Algorithmic Governance and Governance of Algorithms: Legal and Ethical Challenges. Data Science, Machine Intelligence, and Law,* volume 1. Cham: Springer, 2021, 57. Also see Gianclaudio Malgieri and Giovanni Comandé. "Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation" *International Data Privacy Law,* 2017, 7(4): 243–265. https://doi.org/10.1093/idpl/ipx019:243.

[8] Rader *et al.* (n 1).

[9] Jenna Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms", *Big Data & Society*, January–June 2016: 1–12. Compare with Palmiotto's *technical, scientific* and *legal miscodes*. Palmiotto (n 7).

The model aims to construct a link between the data (as input to the algorithm) and the ensuing decisional effects within an implied normative framework (e.g., statute, contract, company bylaws, guidelines etc.). Rather than reflecting the underlying computational processes, the model serves as an abstract template which identifies the requirements that would render data-driven systems more responsive, communicative and engageable from a legal perspective The model is based on the finding that transparency in the sense of explaining "what is" would have little value and even be misguiding or *perfidious*[10] if not complemented with an inquiry about the methodologies and the legal, political and economic justifications underlying the decision.

Contestation on normative grounds both includes whether the inferences made from data are accurate and explainable and whether the decisional rules relying on these inferences are justifiable based on a political, social, contractual, or legal norm or certain moral standard.[11] To put it more concretely, take the example of micro-credit services that heavily rely on ML analysis of mobile phone data of the loan applicant. The deployed ML model, among many other parameters, treats the battery charge level of the phone and the charging frequency as indicative of applicant's lifestyle and thus creditworthiness. In this specific case, the applicant may both contest whether one's failing to charge her/his mobile phone regularly infers a disorganised lifestyle and also whether one's being disorganised is a good reason for refusing her/his loan application.

Taking data as input (akin to the facts in a legal decision) has its own shortcomings in that data collection in ADM is rather selective, that is, the analytic inputs are not simply collected but the input process is rather properly designated as *ingestion*, with all the biological implications of *mastication*, *decomposition*, and *metabolism* that the term entails.[12] Diverse data sources require certain common configurations on differently structured or unstructured digital data—discarding a good deal of information and paving the way to decontextualization.[13] Added to this is the fact that in ML-based ADM, the outcome is an amalgamation of thousands of inferences and parameters, that are not necessarily causal even if correlative.

In sum, contestation requires much more instrumental, abstract and teleological understanding of transparency, reaching beyond explanation and acknowledging that the significance or the harmful character of any decision is definable only by context.[14] In this regard, it becomes clear that biases, omissions, and disparate representation in ADM could not be fully addressed by mere disclosure and openness. The thesis argues that achieving such a goal would require a plethora of regulatory, technical and institutional strategies, systematisations and methodology [15] In terms of operationalizing

[10] Dan L. Burk, "Algorithmic Legal Metrics", *Notre Dame Law Review,* 2021, 96(3):1147-1201, 1166-71.
[11] For sources of norms and the ensuing justifications, see Michele Loi, Andrea Ferrario and Eleonora Viganò, "Transparency as Design Publicity: Explaining and Justifying Inscrutable Algorithms" in *Ethics and Information Technology*, 2021, 23, 253–263.
[12] Dan L. Burk 1158. Also see Louise Amoore and Volha Piotukh, "Life Beyond Big Data: Governing with Little Analytics", Economy and Society, 2015, 44(3): 341-366.
[13] Burk (n 10), 1186.
[14] *ibid*. 1172.
[15] "Algorithmic transparency cannot be understood as a simple dichotomy between a system being "transparent" or "not transparent." Instead, there are many flavors and gradations of transparency that are possible, which may be driven by particular ethical concerns that warrant monitoring of specific aspects of system behavior. [...] Details of the model to disclose might include the features, weights, and type of model used as well as metadata like the date the model was created and its version. A model might also incorporate heuristics, thresholds,

transparency in practice, the systems should be designed and deployed to support contestation, and this implies the implementation of procedures which allow for oversight and which enable scrutiny through algorithmic means.

As such, the *second paper* also reveals the incompatibilities in terms of applying the adversarial method for contesting automated decisions. In this respect, it should be noted that rather than answering, *how to understand the algorithm*, the thesis focuses on the question *how ADM systems should be approached and understood* to identify the mismatches with regard to contestation. Hence the contestation scheme (transparency model) does not primarily solve the transparency problems but aims to inform the legislators, systems developers and operators about how these systems should be designed, configured, deployed, monitored, and documented so that the rules underlying the decision can be challenged. Accordingly, from an individual's perspective who is subject to an automated decision, the thesis does not particularly concretise the technicalities how an automated decision will be contested according to the model but rather identifies the type of elements or dimensions that are necessary to contest the decision on normative grounds.

Further findings of the thesis pertain to the practical implementation of the transparency model, exploring to what extent the relevant provisions in the GDPR could be interpreted in the direction of "contestability". In this part, the thesis develops a systematic and teleological interpretation of Article 22 of the GDPR on automated decisions—focussing on the question how the safeguards to obtain human intervention, express one's views and to contest the decision could practically be implemented. As a result, the *third paper* formulates a typology of transparency impediments, i.e., stemming from technical complexity, economic rivalry and system integrity. This further enables the development of a framework which systemises possible implementation tools and transparency mechanisms under the GDPR as: (1) the *design choices* facilitating interpretability, (2) the *institutional oversight* mechanisms and (3) *algorithmic scrutiny*. The third paper focuses on the specific transparency implications of the "right to contest" as a remedy with a procedural nature. By defining Art 22 as a general provision of due process and the right to contest as the core remedy provided by the GDPR against ADM, the thesis transcends the current debates about the existence and the scope of a so-called "right to an explanation". As a novel approach, the safeguards in Art. 22 para 3 are treated as a different type of obligation distinct from the access and notification rights. Accordingly, while the right to contestation is defined as an *obligation of result,* the notification and disclosure duties (Articles 13 and14) are regarded as *obligations of conduct.*

2.2 Legislative and judicial developments relevant to the findings of the thesis

2.2.1 AI Act proposal

Considering the broadened scope of transparency together with the systemic approach to impediments and requirements (transparency measures) in terms of contesting automated decisions established in

assumptions, rules, or constraints that might be useful to disclose, along with any design rationale for why or how they were chosen." Joshua A. Kroll "Accountability in Computer Systems" in Markus D. Dubber, Frank Pasquale, Sunit Das (eds), *The Oxford Handbook of Ethics of AI*, New York, NY: Oxford University Press, 2020, 180-196, 184.

the *second* and *third* papers, the European Commission's proposed AI Act[16] (released on 21.04.2021) came out as a significantly relevant legislative proposal for the proper implementation of Article 22 of the GDPR and thus, the research question of the thesis.[17] The proposal aims to ensure responsible deployment of AI technologies while addressing the risks for fundamental rights and laying down harmonised transparency rules for certain AI systems. It sets out horizontal rules for the development, commodification and use of AI-driven products, services and systems within the territory of the EU. The draft regulation provides core artificial intelligence rules that apply to all industries.

The proposed Act defines "AI system" as systems that use machine and/or human-based data and inputs to infer how to achieve a given set of human-defined objectives by using *learning, reasoning* or *modelling* (implemented with the techniques and approaches listed in Annex I) and generates recommendations or decisions which influence the environments they interact with.[18] The proposed Act further provides a definition of *general-purpose AI* as systems (irrespective of how they are placed on the market or put into service including open source software) that are intended to perform generally applicable functions such as image and speech recognition, audio and video generation, pattern detection, question answering and translation. A general-purpose AI system may be used in a plurality of contexts and be integrated into various other AI systems. Considering this generic nature of the general-purpose AI, in order to ensure a fair sharing of responsibilities along the AI value chain, such systems are subjected to proportionate and tailored requirements and obligations in a separate title in the Act.[19]

It is made clear that the proposed Act does not affect the obligations of providers and users of AI systems in their role as data controller or processor under the GDPR. Data subjects continue to enjoy all the rights and guarantees awarded to them including the rights related to solely automated individual decision-making by the EU data protection regime. It is mentioned in Recital 58a that the Act should facilitate the effective implementation and enable the exercise of the data subjects' rights and other remedies guaranteed under the personal data protection regime as well as other fundamental rights. Yet, unlike the GDPR, the AI Act proposal does not specify rights and remedies available for individuals affected by the AI systems. Recital 41 also states that compliance with the AI Act does not render the use of the system lawful under other laws of the Union such as the protection of personal data or the use of polygraphs or similar tools to detect the emotional or cognitive state of natural persons. Accordingly, the AI Act proposal does not provide legal grounds for processing of personal data under Article 6 of the GDPR. Considering their scope, the two legislative instruments do not fully overlap. Purely automated decisions subject to Article 22 of the GDPR (which produce legal effects or significantly affect the data subject) are likely to fall within the scope of the general definition of AI systems provided in Article 3 of the proposed Act. Accordingly, the ADM systems which deploy

[16] Proposal for a Regulation of the European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final} (hereafter AI Regulation). The later versions are: the compromised text of the presidency, the consolidated version dated 15.07.2022. https://artificialintelligenceact.eu/wp-content/uploads/2022/06/AIA-FRA-Consolidated-Version-15-June.pdf and the General Approach adopted by the Council of the EU (25 November 2022). https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf. Unless mentioned otherwise references are made to the General Approach dated 25 November 2022.

[17] The proposed Act does not replace but partially overlap with the protections offered by the General Data Protection Regulation (GDPR), as the former's scope is more expansive and not restricted to personal data.

[18] Art. 3(1) and Annex I.

[19] Articles 4a, 4b and 4c (Title I_AGeneral Purpose AI Systems)

techniques and approaches other than those listed in Annex I are excluded. ADM systems subject to Article 22 will also be considered high-risk as they relate to tasks and types of AI use that will be explained in the next paragraph. Furthermore, there are ambiguities regarding the harmonious application of these two legal instruments. For instance, it is not clear how automated decisions falling under the prohibited practices of the proposal will be treated under the GDPR (e.g., *per se* denial of personal data processing on the grounds of legitimate interests of the controller) and whether high-risk systems under the AI Act will also be regarded as high-risk for the purposes of Data Protection Impact Assessment under Article 35 of the GDPR.

The providers who develop AI systems with a view to placing on to the market or putting into service under their own name and the users of these systems (defined as any natural or legal person '*using an AI system under its authority*') are subject to obligations under the Act.[20] The Act defines four different categories of risk, i.e., *unacceptable risk* (prohibited by Article 5), *high-risk* (Article 6), *limited risk* (Article 52) and *minimal risk* (Art. 69).[21] The unacceptable risk (prohibited practices) includes i) subliminal techniques; ii) exploiting vulnerabilities of a specific group of persons due to their age, physical or mental disability; iii) social scoring to evaluate or classify the '*trustworthiness*' of natural persons[22]; iv) real-time remote biometric identification systems in publicly accessible spaces with exceptions for specific law enforcement purposes. The prohibition on subliminal practices and exploitation of vulnerabilities are limited to the cases where such activities are carried out in a manner that causes or is likely to cause the concerned individual or others physical or psychological harm. This formulation is regarded to exclude cumulative harms that occur over time or those caused by other users.[23]

The main regulatory target of the proposed Act is high-risk systems, subject to several compliance requirements including an *ex-ante* conformity assessment combined with strong ex-post enforcement measures. High-risk category includes AI systems that qualify as a product (covered by the legislation listed in Annex II of the proposed Act) and where a third-party conformity assessment is necessary for the placing on the market or putting into service. This also applies to AI systems that are intended to be used as a safety component of a product regulated by the (same) legislation referred to in Annex II. Considering the risks to fundamental rights, Annex III further provides a list of AI systems, i.e., critical infrastructures (e.g., transport) that endanger life and health; biometric ID systems and systems for educational and vocational training, employment, creditworthiness or credit scoring; systems for evaluating the eligibility of natural persons for public assistance benefits and services; systems intended to be used to dispatch, or to establish priority in the dispatching of emergency first response services; systems dealing with migration, asylum applications and border control; and systems intended to assist judicial authority. These systems will be considered high-risk where i) the output of the system is immediately effective with respect to the intended purpose of the system without the need for a human to validate it; or ii) the output of the system consists of information that constitutes the sole basis or is not purely accessory in respect of the relevant human action or decision which may lead to a

[20] This further extends to importers and distributors of these systems. For concerns that the scope of the Act may be unfeasibly wide, see Lilian Edwards, "The EU AI Act proposal: a summary of its significance and scope" Ada Lovelace Institute, April 2022. https://www.adalovelaceinstitute.org/ resource/eu-ai-act-explainer/

[21] Ida Varošanec, "On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI" (2022) *International Review of Law, Computers & Technology*, 36:2, 95-117, 101.

[22] The provision does not provide clarity, for instance, whether a system which assesses families for the risk of child neglect or abuse would be covered.

[23] Michael Veale and Frederik J. Zuiderveen Borgesius, "Demystifying the Draft EU Artificial Intelligence Act – Analysing the Good, the bad, and the Unclear Elements of the Proposed Approach." (2021) *Computer Law Review International* 22 (4): 97–112.

significant risk to the health, safety or fundamental rights. As such, in comparison to the requirement of "solely automated processing" in Article 22 of the GDPR, the AI Act covers a wider range of automated decisions by only excluding the AI systems that are purely accessory.[24]

Risk assessment required by the proposed Act not only concerns the function performed by the AI tool, but also the specific purpose and modalities used by the system— together with the extent of potential harms and whether the impacted persons are dependent on the outcome produced by the AI system or in a vulnerable or weak position. The Act puts special emphasis on the concept of "intended purpose" which is designated as a type of information which should be included in the instructions accompanying high-risk AI systems. The intended purpose also includes the specific geographical, behavioural or functional setting within which the high-risk AI system will be used. Hence, it may be expected that the modalities, procedures and practices that will be developed for the risk assessment of the intended uses and possible misuses under the proposed AI Act will contribute to the interpretation and efficient application of the purpose limitation principle under the GDPR.[25] The proposed Act further takes into account whether the current EU legislation provides effective measures to minimise risks and redress damages (together with the possible reversibility of the decisions produced by an AI system). Whether the impacted person is dependent on the outcome produced by an AI system, be it for practical or legal reasons, and the possible vulnerable position in relation to the user of an AI system are also of consequence in the assessment of risk. As such, these risk assessment factors could provide interpretative guidance about the "legal or similarly significantly affects" as prescribed by Article 22 of the GDPR.

The proposal also contains a specific provision requiring that training, validation and testing data shall be subject to appropriate data governance and management practices. The provision includes references to the relevant design choices, data collection processes, data preparation operations (e.g., annotation, labelling, cleaning, enrichment and aggregation), the relevant assumptions (based on the information that the data are supposed to measure and represent) and the geographical, behavioural or functional characteristics of the domain.[26] Together with these, further specific requirements regarding the technical documentation, record keeping and the characteristics, capabilities and limitations of performance of the high-risk AI systems correspond with the necessary information as defined in the transparency model laid out in the *second paper* (e.g., the input data, context, impact of the decision and the decisional rules). In sum, the relevant provisions of the proposed Act may be regarded as elucidatory in relation to the transparency mechanisms for the implementation of Article 22 of the GDPR as explained in the *third paper*.

Of significant relevance to Article 22 of the GDPR, the proposal puts special emphasis on human oversight in case of high-risk AI systems. '*Human agency and oversight*' are the core principles of ethical AI which come into practice as '*human in the loop'* referring to the capability of human intervention and

[24] See above Ch.4, sec. 2.1 *Decisions based solely on automated processing, with legal or similarly significant effects*.
[25] In the initial version of the Act (21 April 2021), the risk assessment also required a consideration of *reasonably foreseeable misuse*, where misuse is defined as use for another than the intended purpose (see Articles 9(2b) and 13(3) of the proposed AI Act). Yet, the issue is still undecided since Art. 9 and 13 of General Approach of the Council (see above n 12) omits the references to *reasonably foreseeable misuse.*
[26] Art.10 of the AI Act proposal. Yet, it should be noted that the efficiency of the provision is to a certain extent diminished as Article 42(1) states that AI systems that have been trained and tested on data which reflect the specific geographical, behavioural or functional setting within which the system is intended to be used shall be presumed to be in compliance with the data governance requirements.

'*human on the loop'* as the capability to oversee the overall activity.[27] Article 14 (as it currently stands) of the proposed AI Act sets a general obligation for high-risk AI systems that they should be effectively overseen by natural persons. The relevant provision in the AI Act addresses both the developers and users of AI systems. Developers shall implement measures before the high-risk AI system is placed on the market or put into service and further identify measures that are appropriate to be implemented by the user. High-risk AI systems shall be designed in a way that will enable the human overseer to understand the capacities and limitations of the system, duly monitor its operation and remain aware of potential automation biases stemming from the possible tendency of relying or over-relying on the output produced by a high-risk system. The human overseer should also be able to correctly interpret the high-risk AI system's output to decide, in any particular situation, not to use the high-risk AI system or otherwise override its output. The proposed Act is not clear whether and where humans shall have the final word on the decision. Nevertheless, Article 14 is a detailed provision which elucidates many issues relating to human oversight in ADM. Here again, mechanisms and methodologies that will be developed for compliance with the human oversight requirement could be expected to assist the implementation of human intervention safeguard under Article 22 of the GDPR.

Overall, the wide span of requirements and obligations in the AI Act proposal relating to the development and deployment of high-risk AI systems clearly align with the theoretical underpinnings of this thesis which primarily approaches contestation of ADM as a procedural matter.[28] It is particularly of significance that the proposed Act clarifies that the humans tasked with oversight must have an understanding of both the capacities and the limitations of the systems and thus those parts of an artefact that one may not know through disclosure and access. The proposal provides substantial support to distinguish and crystallise the matters of transparency and contestation in a procedural context and this is further confirmed by Recital 38 which states that where AI systems are not sufficiently transparent, explainable and documented*, procedural fundamental rights*, such as the right to an *effective remedy* and *fair trial* as well as the *presumption of innocenc*e could be hampered. The proposed legislation further acknowledges that designing IT systems implies evaluative choices such as the descriptive features of the data analysed or the set of assumptions with respect to the information that the data are supposed to measure and represent. The proposal is a significant leap forward for the legibility of ADM systems owing to its diverse conception of transparency—recognizing distinct forms of opacity inherent to ADM systems, which is vital in developing (technical and nontechnical) solutions to respond to a multitude of transparency demands. Yet, despite important transparency measures, the Act does not properly address the possible conflicts with the IP rights.[29] The proposal may also be criticized for failing to take into account the power imbalances between those who develop and deploy the technology, and those who are subject to its decisions.[30]

---

[27] High-Level Expert Group on Artificial Intelligence, the EC Ethics Guidelines for Trustworthy AI, 2019 https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html. Also *see*, White paper on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[28] As elaborated in the *2nd* and *3rd* paper of the thesis, such approach to ADM will also help calibrate the legal scrutiny according to the legal domain and the perceived risks and in particular, the different review criteria in public and private sector decision-making.

[29] Varošanec, (n 21), 106.

[30] Sebastian Klovig Skelton, "Europe's Proposed AI Regulation Falls Short on Protecting Rights." 2021. Computer Weekly.com. https://www.computerweekly.com/feature/Europes-proposed-AI-regulation-falls-short-on-protecting-rights. For more views on the Proposal arguing that it does not provide an effective framework for the enforcement of legal rights and duties, failing to ensure legal certainty and consistency, see, Nathalie Smuha and others, "How the EU can achieve legally trustworthy AI: a response to the European Commission's

2.2.2 Digital Services Act

Another important legislative development relevant to contestation and transparency is the Digital Services Ac (DSA)[31] which primarily concerns online intermediaries and platforms such as online marketplaces, social networks, content-sharing platforms, app stores, and online travel and accommodation platforms. The main objectives of the Act are to improve the protection of consumers and their fundamental rights online and establish a powerful transparency and accountability framework for online platforms with a view to foster innovation, growth and competitiveness within the single market. The DSA introduces various transparency and reporting obligations for different types of actors, i.e., *intermediary services, hosting services, online platforms* and *very large online platforms* (VLOPs)[32] as defined by the Act.

Article 15(1) reads as: "providers of intermediary services shall make publicly available, in a machine-readable format and in an easily accessible manner, at least once a year, clear, easily comprehensible reports on any content moderation that they engaged in during the relevant period." Under Article 15(1)(c), providers of intermediary services shall further provide meaningful and comprehensible information about their content moderation activities including the use of automated tools.[33] Any use of automated means for the purpose of content moderation, a specification of the precise purposes, indicators of the accuracy and the possible rate of error and any safeguards applied are also among the information that are included under the transparency reporting obligations (Art. 15 (1)(e)). Of relevance to contestation, Article 14(1) stipulates that intermediary services should incorporate information about content moderation measures and tools (including algorithmic decision-making) in their terms and conditions. Recital 70 elaborates that recipients of the service should be appropriately informed about how algorithms impact and influence the way information is displayed.

In Article 27, further transparency obligations are brought for online platforms that use fully or partially automated recommender systems to suggest or prioritise specific information. The main parameters of the system, including the criteria which are most significant in determining the information suggested and the reasons for the relative importance of those parameters, shall be set out in plain and intelligible language in the terms and conditions of the service.

---

proposal for an Artificial Intelligence Act" (2021) *Artificial Intelligence - Law, Policy, & Ethics eJournal*. https://ssrn.com/abstract=3899991 or http://dx.doi.org/10.2139/ssrn.3899991

[31] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act). For a general evaluation of the Act, *see* Joris van Hoboken, João Pedro Quintais, Naomi Appelman, Ronan Fahy, Ilaria Buri and Marlene Straub, "Putting the Digital Services Act Into Practice: Enforcement, Access to Justice, and Global Implications" (March 10, 2023). Amsterdam Law School Research Paper No. 13, 2023, Institute for Information Law Research Paper No. 03, 2023, Verfassungsbooks. https://ssrn.com/abstract=4384266

[32] *See* Art. 33. For *very large online platforms* (VLOPs) or *very large online search engines* (VLOSEs) status, see the designating decision on 25 April 2023.
https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413

[33] DSA Art. 20(6) also provides that certain decisions shall not be taken by solely automated means. Also see Art. 17(3)(c).

In relation to advertising on online platforms, the online interface should facilitate recipients of the service to easily access relevant and meaningful information, in real time. This information should pertain to the primary parameters utilized to determine the recipient and provide guidance about modifying those parameters (Article 26(1)(d)).

VLOPs and VLOSEs are under heavier obligations to identify, analyse and assess any systemic risks stemming from the design or functioning of their service and its related systems, including algorithmic systems. In their risk assessment, these providers shall take into account the factors such as the design of their recommender systems and any other relevant algorithmic systems. This risk assessment shall be carried out at least once every year and in any event prior to the deployment of functionalities that are likely to have a critical impact on the risks.[34] In the context of online advertising, VLOPs have an obligation to compile a repository and make it publicly available. The repository shall be accessible through a searchable and reliable tool that allows multicriteria queries through application programming interfaces (APIs) (Art. 39(1)). The repository shall contain information on "whether the advertisement was intended to be presented specifically to one or more particular groups of recipients of the service and if so, the main parameters used for that purpose including where applicable the main parameters used to exclude one or more of such particular groups" (Art. 39(2)(e)).

The DSA also refers to certain transparency measures and implementation tools, most importantly the external and independent auditing to ensure compliance of VLOPs with the obligations set forth in the Act. This includes any additional commitments made through codes of conduct and crisis protocols, where applicable. Recital 92 provides that VLOSEs should provide the necessary cooperation and assistance to the organisations carrying out the audits. This will include giving the auditor access to all relevant data and premises necessary to perform the audit properly, including, where appropriate, to data related to algorithmic systems. The Act also addresses certain secrecy concerns in respect of the information obtained from the providers of VLOPs and third parties in the context of the audits. In Article 37(2), it is also made clear that those secrecy and concealment shall be kept at the minimum and should not adversely affect the performance of the audits and other provisions of this Regulation— in particular those on transparency, supervision and enforcement. This highlights the significance of employing a diverse set of transparency tools to address the impediments stemming from IP rights, without exceedingly compromising the contestability. Under Article 40, VLOPs shall provide the Digital Services Coordinator of establishment or the Commission with access to data that are necessary to monitor and assess compliance with this Regulation. They are also obliged to explain the design, the logic, the functioning and the testing of their algorithmic systems, including their recommender systems.

Together with the internal complaint-handling system in Article 20, the provisions of the DSA may be regarded to as a form of due process for downranking, providing for duties to give reasons and to hear

[34] Recital 84 of the DSA reads as "When assessing such systemic risks, providers of very large online platforms and of very large online search engines should focus on the systems or other elements that may contribute to the risks, including all the algorithmic systems that may be relevant, in particular their recommender systems and advertising systems, paying attention to the related data collection and use practices." Recital 85 further provides that all supporting documents relating to the risk assessments, such as underlying data and data on the testing of their algorithmic systems, shall be preserved so that subsequent risk assessments could build on each other and show the evolution of the risks identified.

appeals.[35] Even though not explicitly referring to *input data* or *decisional rules* as prescribed in Chapter 3, the disclosure of information about the parameters and the rationale behind their relative weight in a specific decision may help formulate a contestation or scrutiny— enabling normative challenges in a manner similar to the transparency model presented in the thesis. Yet, it should be borne in mind that as none of the requirements or the elements of the model is explicitly specified in the DMA, even a partial application of the model would require a purposeful and broad interpretation of the relevant provisions.

2.2.3 Digital Markets Act

The other EU legislation, Digital Markets Act (DMA)[36], aims to ensure that online platforms which act as "gatekeepers"[37] in digital markets behave in a fair way. Together with the Digital Services Act, DMA is one of the centrepieces of the European digital strategy. DMA specifically classifies online advertising as a type of core platform service, including any advertising networks, advertising exchanges and any other advertising intermediation services. In relation to ADM, DMA includes provisions directly related to opacity in online advertising and the measurement of its effectiveness.

According to Article 6(8), advertisers, publishers, or their authorised third parties will be given access to: performance measuring tools of the gatekeeper; and the data necessary to carry out independent ad verification. Such data shall be provided in a manner that enables advertisers and publishers to run their own verification and measurement tools to assess the performance of the core platform services provided by the gatekeepers.

Yet, the provisions of the DMA do not intend to resolve the issues arising out of behavioural profiling of the end users or the ensuing automated decisions. The transparency provisions in the DMA are not designed to directly address the adverse effects or due process violations of automated decisions in online advertising.[38] "Contestability" within the context of DMA rather refers to the ability of undertakings to effectively overcome barriers to entry and expansion and challenge the gatekeeper on the merits of their products and services. As such, contestability under the Act, generally relates to competitive dynamics and market regulation.[39]

DMA approaches profiling in online advertising from the perspective of fair and contestable markets. In this respect, as Recital 72 states: "the data protection and privacy interests of end users are relevant

---

[35] Paddy Leerssen, "Algorithm Centrism in the DSA's Regulation of Recommender Systems", Verfassungsblog, 22.03.2022, https://verfassungsblog.de/roa-algorithm-centrism-in-the-dsa/

[36] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)

[37] The platforms that qualify as "gatekeeper" are the digital platforms that provide an important gateway between business users and consumers – whose position can grant them the power to act as a private rule maker and thus creating a bottleneck in the digital economy.
economy. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_6423

[38] However, as mentioned by Micova, "[i]f effectively implemented and with fairness toward users as a priority, the transparency provisions may help to encourage a move away from the more invasive targeting techniques by giving more visibility to the effectiveness and relative value of contextual, broadly segmented, and other types of advertising." Sally Broughton Micova, "DMA:Transparency Requirements in Relation to Advertising", Issue Paper, November 2022 CERRE project 'Effective and Proportionate Implementation of the DMA' https://cerre.eu/wp-content/uploads/2022/11/DMA_TransparencyRequirementsinAdvertising.pdf

[39] Recital 32.

to any assessment of potential negative effects of the observed practice of gatekeepers to collect and accumulate large amounts of data from end users." The Recital provides that an adequate level of transparency of profiling practices employed by gatekeepers —including, but not limited to, profiling within the meaning of Article 4(4) of the GDPR— would facilitate contestability of the core platform services. Accordingly, gatekeepers should at least provide an independently *audited description* of the basis upon which the profiling is performed, including the processing operations and the purpose for which the profile is prepared and eventually used.[40] Under Article 15, the Commission may further adopt an implementing act to develop the methodology and procedure of the audit. It is made clear that while making the audited description available to public, the gatekeeper may take into account of the need to protect its business secrets.

The transparency provisions of the DMA do not envisage an individual contestation of profiling-based decisions and therefore, the transparency model of the thesis could have limited application within the context of the Act. It is important to note that the DMA stands out from other legislations as it gives priority to facilitating access to data for the wider goal of promoting contestable and fair markets. According to the Act, the objective of improving contestability and fairness within the advertising ecosystem will only be served if most transaction data is shared. What seems rather of significance for the thesis is the specific refence that links contestability with the results of the external audit (together with the provisions for the development of the relevant methodology and auditing procedures).[41]

2.2.4 The Case C-817/19, Ligue des droits humains

A significant judicial development is the ruling of the Grand Chamber of the European Court of Justice (ECJ) on 21.06.2022 (C-817/19, *Ligue des droits humains (LDH)*) which pertains to the processing of passenger data under the PNR Directive[42]. The judgment provides insights that are also relevant for the overall approach of the thesis towards ADM.[43]

The PNR Directive provides a series provisions relating to the automated processing of PNR data together with certain safeguards, though without granting individual rights for contestation. Article 6 of the Directive permits the processing of PNR data against pre-determined criteria to identify persons who require further examination by the competent authorities, based on the risk of involvement in a terrorist offence or serious crime. The same Article also contains safeguards providing that the automated processing must be "carried out in a non-discriminatory manner" and the pre-determined criteria should be targeted, proportionate, specific, and open to review by the competent authorities.

[40] Recital 72 and Article 15 of DMA.
[41] (Art. 15(2). Algorithmic scrutiny and audit are among the transparency tools and mechanisms examined in Ch.4 sec.5.4.
[42] Directive (EU) 2016/681 of the European Parliament and of the Council of 27 April 2016 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime.
[43] The ECJ decision also evaluates various aspects of the Directive EU 2016/680 (Police Directive). As explained in the Introduction (Ch.1, sec.3), Police Directive has been left out of the scope of the thesis for the reason that the relevant Article 11 does not allow the data subject to express her/his point of view and contest the decision. As such the applicability of the transparency model of the thesis under the Police Directive depends on the question whether a *right to contest* could be implied from the *right to obtain human intervention* in Article 11(1) or from the general rules of procedure. On this matter, it could simply be mentioned that if the involvement of human intervention or review by non-automated means is transparent and not concealed from the data subject, the decision may inevitably be subject to contestation at least on the grounds of arbitrariness and malintent.

The retention and processing of PNR data revealing a person's race or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, health, sexual life, or sexual orientation is prohibited. Any positive match resulting from the automated processing of PNR data should be individually reviewed by non-automated means.

In the *LDH* case, the Court was asked about the compatibility of the PNR Directive with the EU Charter of Fundamental Rights (the Charter). Among various issues, the Court provides certain clarifications on how the advance assessment of PNR data by automated means could be organised in conformity with the Charter. Regarding the automated processing based on pre-determined criteria (Question 6), the Advocate General (AG), in his Opinion, asserts that the algorithms used for the analysis must function transparently, and the result of their application must be traceable. As such, in line with the approach of the thesis, the AG confirms that the requirement of transparency cannot be understood as the disclosure of the profiles to the public but rather the traceability of the result. According to the AG, the safeguard of individual review by non-automated means must enable an understanding of why the program arrived at a specific match. Transparency in the functioning of the algorithms is a necessary precondition for the data subjects to be able to exercise their rights to complain and their right to an effective judicial remedy. It is of paramount importance that (both *ex ante* and *ex post)* supervision by an independent authority should be able cover all aspects of the automated processing of PNR data, including the selection of the databases used for comparison and the pre-determined criteria.[44] As the Opinion illustrates, an interpretation of the PNR Directive compatible with the Charter may relate to many requirements which resonate with the transparency model of the thesis.[45] In that sense, the interpretation provided in the Opinion, especially the reference to database selection and the reasons for specific match, clarifies that the transparency model of the thesis could be applicable while conducting the scrutiny of the PNR system. An important point in the Advocate General's Opinion (which is also confirmed by the Court) is that the requirement of pre-determined criteria in Article 6(3)(b) is understood as precluding "the use of artificial intelligence technology in self-learning systems ('machine learning'), capable of modifying without human intervention or review the assessment process."[46]

The Court further rules that where it is not possible to understand the reasons why a given program arrived at a positive match, the data subjects would be deprived of their right to an effective judicial remedy enshrined in Article 47 of the Charter.[47] This proves the importance of developing a multi-dimensional and versatile approach to transparency—including the technical, procedural, administrative and institutional measures which could enable the review of the systems that are not accessible by humans or through mere disclosure.[48] As mentioned in Chapter 4 above, determining the optimal combination of these transparency measures is a context-specific task that needs to take into

[44] Opinion of Advocate General Pitruzzella, Case C-817/19, *Ligue des droits humains v. Conseil des ministres,* delivered on 27 January 2022. para. 228.

[45] For instance, the requirements of the thesis' transparency model regarding how data and data features have been collected and selected may help overcome the risks resulting from the ambiguity of the PNR Directive as to the relevant databases. Information on how data and data features have been collected and selected stands as an essential step of transparency for contestation which may be applicable for a wide range of automated decisions.

[46] C-817/19, *Ligue des droits humains,* para. 194. It should be noted that such understanding may not be truly accurate since not all ML systems learn from new data as it becomes available. Static systems depend on human intervention for update. *See* Ch.1, sec.1, note.14.

[47] *ibid.,* para. 195.

[48] See above Ch.4, sec. 5.3

account the technical limits, possible gaming-strategies, and competition-related concerns.[49] After all, the prohibition of a system could only be justified where all the feasible transparency mechanisms and implementation tools prove to be insufficient.

2.2.5 The emerging regulatory landscape

The newly enacted and the upcoming legislative agenda of the European Commission under the 2020 Data Strategy and the AI initiatives, lay out a regime for data use/transactions and AI deployment, accompanied with further sectoral adjustments. This emerging regulatory landscape set forth by the AI Act, DSA, DMA, DA and DGA provide a wide range of obligations that aim to ensure transparency and enable scrutiny of ADM in various contexts and dimensions.[50] Furthermore, we see an elaboration and deployment of various implementation tools, mechanisms, methodologies and technical and institutional measures, e.g., algorithmic audit, APIs for data access, code of conduct, standards and independent oversight. The development and refinement of these transparency measures and mechanisms will eventually improve the general arsenal of transparency tools for contesting automated decisions. The advances in this front is expected to facilitate the application of the transparency model of the thesis through the implementation tools as defined and systemised in Chapter 4. Overall, the increased data access and the regulatory tightening in the field of AI and ML-based applications will create a legal landscape with better remedies for transparency and for the legibility of the normative configuration of ADM systems under scrutiny.

Of specific relevance to the transparency model of the thesis, the overall ADM framework relating to transparency and contestation in the EU *acquis* acknowledges that access to and understanding of data are equally important as algorithmic legibility. As the implementation of this emerging regulatory framework progresses, it could be expected that the approach to data in ADM as *decisional input* (in Chapter 3) would have a wider applicability for a range of automated decisions. Having said that, it should be noted that the emphasis of the transparency model of the thesis on the *decisional rules* and the *context* of ADM cannot be regarded to have reached general acceptance as of 2023.

More importantly, these significant legal and policy innovations introduced by the emerging legal framework (which require new processes and methodologies to ensure their efficiency) are scattered throughout a long and intricate web of Recitals and Articles.[51] In many cases, it could be observed that each of these legal formulations lean towards or prioritise one or more of the dimensions explained in the transparency model of the thesis. This confirms the need for a methodological and systemic approach to transparency in the context of ADM to guide the implementation process of various provisions under the EU *acquis*. In this respect, the transparency model and the systemisation of the implementation tools offered by the thesis may also be seen as a preliminary structure for developing a comprehensive framework to map the legal territory.

---

[49] See above Ch.4, sec.6.

[50] How the DA (proposal) and the DGA contributes to this framework has been explained in Ch.5, sec.3.5

[51] Laura Edelson, Inge Graef, and Filippo Lancieri,"Access to Data and Algorithms: For an Effective DMA and DSA Implementation" (CERRE, March 2023), https://cerre.eu/publications/access-to-data-and-algorithms-for-an-effective-dma-and-dsaimplementation

2.3 Contributions of the thesis and the questions yet to explore

The findings of the first three papers of the thesis specifically focussing on the reasons, modalities and mechanisms of transparency in ADM ultimately bring about the below conclusions and contributions while revealing certain questions which are yet to be explored.

Regarding algorithmic transparency, the first take-away of the thesis is that the contestation of ML-based decisions is not about reading the computer code, but rather relates to the question how these systems make up the regulatory realm that we are subjected to. Approaching these systems as procedural mechanisms with implicit 'normativity', the thesis supplies both the conceptual and the terminological arsenal to encompass various dimensions of ADM within the context of contestation. The procedural approach to transparency and contestation enables a model which is abstract enough to accommodate different motivations and legal grounds underlying the use of ADM both by public administration and private companies in a multitude of contexts.

Reaching beyond the current debates shaping around the concepts like fairness, equality or non-discrimination, the thesis offers a theoretical vantage point by approaching ADM systems as procedural mechanisms, results of which could be challenged on normative grounds. It is the basic tenet of this study that these value-laden, domain-dependent and quasi-legal concepts can only address a fragment of the problem and thus, cannot serve as a theoretical basis for a general approach to the problem of contestation.[52] The approach of the thesis to ADM also enables a more granular regulatory perspective which—rather than revolutionizing an upheaval and rewriting of the legal realm— regards contestation also as a matter of a transformation and adaptation of the procedural tools and mechanisms that are already in place in modern legal systems.

The thesis illustrates in many facets the significance of interdisciplinary efforts for theorisation and model building together with the ensuing typologies and conceptual frameworks as the primary methods to break down complex, multifaceted phenomena such as ADM into manageable parts. Though categories, typologies or taxonomies may not seem sufficiently precise, the focus should not be on the overlaps between the boundaries but rather on the fact that structuring enables a macro-view framework to analyse how different dimensions of the problem relate to each other. When applied to a specific ADM problem, the typologies and models should be understood within the relevant legal, economic and social context and not simply as analogues of abstract compartmentalisations.

Like any system, ADM can be viewed from a number of different perspectives, which may yield different types of decompositions of the system.[53] The relation that a model constructs with the object

---

[52] As Radbruch puts: "[w]hile justice directs us to treat equals equally, unequals unequally, it does not tell us anything about the viewpoint from which they are to be deemed equals or unequals in the first place; moreover, it determines solely the relation, and not the kind, of the treatment." Gustav Radbruch, "Legal Philosophy" in *The Legal Philosophies of Lask, Radbruch and Dabin* (Edwin W. Patterson ed., Kurt Wilk trans.) Cambridge, Massachusetts: Harvard University Press, 1950, 10. Kroll also argues "[...]although many (or all) stakeholders in a particular context may wish an AI system to behave fairly, what is fair for some may not be fair for others. Setting out rules for what constitutes fairness must, of its nature, set these stakeholders in tension with each other." Kroll (n.15), 186. Also see Walter Bryce Gallie, "Essentially Contested Concepts," in *Proceedings of the Aristotelian Society,* 1955, 56:167–198.

[53] Stuart. A.Kauffman, "Articulation of parts explanations in biology" in R. C. Buck and R. S. Cohen (Eds.), *Boston Studies in the philosophy of science,* 1971, Vol. 8, 257–272. Also see, William Wimsatt, *Re-engineering*

of interest heavily depends on what the model is intended to be used for. Hence, the contestation model offered by the thesis should be seen as an attempt to systemise/typologise what we ought to know about ADM systems to contest their outcome. Although, it is inevitable that a model involves various deliberate abstractions, a theoretical foundation and the derived model is the only way to unpack a "multiple decomposable" system (entity).[54] In this respect, the methodological approach and systemisations developed throughout the thesis are vital to navigate the complex regulatory landscape of algorithmic transparency. As a further contribution, the typologies, systemisations, theoretical frameworks and models developed throughout the thesis may be seen as a part of a wider legal lexicon-building effort which contributes to the development of an all-encompassing contestation framework— serving both as a guidance for the design and the audit of ADM systems, and also as a *scheme* for the ex-post scrutiny of specific decisions on several grounds and against different actors.

Regarding matters that deserve further research efforts, it could be mentioned that despite the legal affordances explained throughout the papers, there remain many gaps to be bridged between the right to contest as provided in the GDPR and its practical application. Hence, as concluded in the third paper, the transparency requirements and obligations scattered around various legal instruments would still need to be developed into a contestation scheme— encompassing various grounds that a specific case of contestation could be based on as well as the diversity of the actors which could be held liable. The crux of the matter is determining the optimum extent of transparency and the appropriate mode of implementation, without prejudice to the integrity of the systems or the legitimate interests of the stakeholders. Considering the remaining uncertainties and open questions related to the feasibility and the efficacy of possible technical and legal solutions, the thesis could serve as a launching pad for further legal research which will elaborate on the limits and the impediments to human-intelligible models. That is, where there are genuine technical and legal barriers and where complexity and/or legal claims are used as a pretext for unsubstantiated or unlawful secrecy practices. Such inquiry should consider that even if it is not feasible or preferable to disclose, in an explicit sense, what algorithms do and which inaccuracies they might have, it could still be possible at least to embody them to such a degree that the users know when to rely on their results and when to become distrustful without compromising their predictive power.[55] In this respect, the regulatory options prescribed in the third paper provide a preliminary layout to examine how different transparency mechanisms can be implemented in a coherent framework. Accordingly, there is also more work to be done on the question of how design (as a technical process by which a set of specifications are translated into computer code) could be integrated with normative prescriptions, algorithmic tools and institutional mechanisms aiming to enable effective contestation.[56] This line of research should also consider that

---

Philosophy for Limited Beings: Piecewise Approximations to Reality, Cambridge, Mass: Harvard University Press, 2007, 181.

[54] William Wimsatt, "Complexity and organization" in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Vol. 20, 67–86). (1972). Dordrecht: D. Reidel

[55] For this approach also referred to as "practical transparency", see Johannes Paßmann, and Asher Boersma, "Unknowing Algorithms: On Transparency of Unopenable Black Boxes" in Mirko Tobias Schäfer, Karin van Es (eds), *The Datafied Society: Studying Culture through Data* (Amsterdam: Amsterdam University Press 2017), 139–146.

[56] "Design refers both to the code-based architecture of a technological system and the social process by which corporate interests, legal requirements and other mandates are translated into that code. It captures both the technical and organizational elements that impact how technologies are made and what they do for (and to) their users." Ari Ezra Waldman, *Industry Unbound: The inside Story of Privacy, Data, and Corporate Power*, Cambridge University Press, 2021, 164. Also *see* Karen Yeung, "Hypernudge: big data as a mode of regulation by design" *Inf Commun Soc,* 017, 20:118; M. Ryan Calo, "Digital market manipulation" *George Washington Law Review* (2013), 82:995.

transparency in the form of disclosure may be deployed as a long-term strategy to avoid liability (e.g., for poor accuracy or harmful consequences) and more importantly, models may be rendered human-intelligible to purposely manipulate the behaviour toward some desired outcome.[57]

3. On the IP rights as impediments

3.1 Summary of findings and contributions

As the most important legal framework which could impede transparency efforts, the final part of the thesis on IP rights provides a macro-view of the potential areas of conflict between the transparency requirements and the relevant IP regimes—i.e., copyright, sui generis database right and trade secret protection. The *fourth* paper brings about two essential findings: *First*, *conventional* IP rights (as subject to a set of mandatory exceptions and limitations) do not generally give rise to a significant conflict with the transparency requirements developed in the *second paper*. It is the rights that address digital technologies (i.e., software protection, anti-circumvention rules[58], sui generis database rights) which mainly create barriers in the implementation of the transparency measures.[59] *Second*, the analysis further reveals that the transparency measures (i.e., *design choices* facilitating interpretability, the *institutional oversight* mechanisms and the *algorithmic scrutiny)* that go beyond access and disclosure are more likely to create conflicts with the IP rights. The type of use envisaged by these measures do not easily fit in the exceptions and limitations provided by the IP regime. [60]

In terms of addressing the implications of IP rights as counter-arguments to transparency demands in ADM, as a methodological contribution of the thesis, the analysis in the *fourth* paper is not based on IP types but rather on a taxonomy of ML as expressional and utilitarian (functional) elements. ML elements classified in this way lay out the necessary basis to explore how the implementation of transparency measures interacting with these elements might give rise to claims under copyright law, sui generis database right, patent law, and trade secret law. Along with this, the distinction of ML data into the categories of *training*, *actual* and *output* data for the purpose of identifying IP relevance is another contribution of the thesis. This way of approaching ML data offers a methodology which enables the analysis of legal implications of different types of data use or access requirements.

---

[57] Burk (n 10), 1191.
[58] Art.6 of the Infosoc Directive.
[59] As Gervais notes, the expansion of "primary IP rights" through "secondary rights" have eroded exceptions and limitations as a key instrument in calibrating IP rights and reconciling with other public interests. He refers to "*primary* IP rights" as copyright, trademark, design and patent law that have been established by the international treaties and implemented in many national laws for well over a century. This parallels with the *conventional IP rights* as referred to in the fourth paper of the thesis. As "*secondary* IP rights", Gervais refers to the rights which expands the primary rights with a view to address new technological developments. Daniel J. Gervais, "Introduction to the future of intellectual property" in Daniel J. Gervais (ed,) *The Future of Intellectual Property*, Cheltenham, UK; Northampton, Massachusetts: Edward Elgar, 2021, 1-7, 1. The expansion of primary rights has come along with certain marginalization of the principles of morality and *ordre public*, narrowing the scope of IP policymaking. Sigrid Sterckx and Julian Cockbain, *Exclusions from Patentability: How Far Has the European Patent Office Eroded Boundaries?* Cambridge: Cambridge University Press, 2015. Also see Aisling McMahon, "Gene Patents and the Marginalisation of Ethical Issues" (2019), *European Intellectual Property Review* 41(10):608–620.
[60] Paul de Laat, "Algorithmic decision-making employing profiling: will trade secrecy protection render the right to explanation toothless?", *Ethics and Information Technology* 24(2), June 2022, 17. https://doi.org/10.1007/s10676-022-09642-1.

The thesis further contributes by improving the conceptual clarity as to the nebulous terms of *algorithm* and *ML model,* as there exists a multitude of perspectives, narratives and vocabularies which deploy these terms inconsistently, imprecisely and interchangeably. Only after the establishment of the conceptual relation between these terms, it becomes possible to explain why both ML models and algorithms as abstract formulas cannot amount to a solid expression eligible to copyright protection.[61]

Among the different types of protection analysed in the *fourth* paper, the application of trade secret (TS) rules to ML systems and data distinguishes as a more controversial and difficult topic. Because copyright and patent law in many respects fall short of protecting ADM systems against legal demands for access and disclosure, TS protection increasingly becomes the preferred type of legal remedy for those who control data or ADM systems. Therefore, more heated debates may be expected regarding the application and suitability of the TS protection to ML elements. The "blanket" cover provided by TS protection raises concern in that excessive reliance on trade secrets could give rise to disincentives for potential investors as it creates uncertainties with regard to the validity and enforceability of data transactions. This contradicts the policy initiatives and the legislative agenda of the European Commission which aims to increase the availability and accessibility of data while ensuring the accountable use of AI systems. That being said, despite the extensive set of obligations and requirements to ensure transparency and accountability in ADM, the proposed AI Act offers very little guidance for a potential conflict with the IP rights especially where the disclosure, access and processing of data as a part of an AI system would jeopardise the competitive interests of the system owner or render the system prone to manipulation. The Explanatory Memorandum of the proposed AI Act merely asserts that the increased transparency obligations would not disproportionately affect the right to intellectual property just because the obligations and requirements in the AI Act are limited to the minimum information necessary for supervision and enforcement authorities. In a similar fashion, despite the several references to the TS Directive in the recent Data Act proposal of the Commission (23.02.2022), both legislative proposals do not provide any guidance about how the provisions relating to access or sharing of data will be applied to cases where the data in question contains or constitutes a trade secret. This lack of clarity confirms the importance of alternative solutions that will be highlighted in the below section.

3.2 Conflict of fundamental rights and further solutions

What remains untouched in the thesis is a further proportionality analysis as a way to resolve the conflict between the IP and personal data protection as fundamental rights. The part of the thesis on IP rights is confined to a *prima facie* analysis aiming to draw a macro picture of possible IP conflicts and thus, does not extend to the question how the principle of proportionality could be applied to the context of contestation. The tension between the protection of the right to intellectual property and other fundamental rights manifests itself on various points relating to the interpretation of the exclusive rights, the scope of the exceptions and limitations and enforcement. In resolving these conflicts, the CJEU employs a fair balance test based on the proportionality principle provided in Article 52(1) of the Charter of Fundamental Rights of the European Union (the Charter).

[61] This is leaving aside the possible expression or representation of algorithms in tangible forms such computer program in programming language. See above the Fourth Paper, section 4.2 *Copyright protection as creative expression* and fn. 100.

The matter has come before the European Court in various copyright injunction cases brought against those who offer intermediary services. The early rulings (e.g. *Promusicae* and *Sabam)*[62] have been subject to intense criticism due to the vagueness and the ensuing inconsistency in the application of the principle of proportionality.[63] The Court's decisions are generally found to be creating legal uncertainty as they lack concrete guidelines for the required balancing between fundamental rights.[64] Critical voices argue that the existing case-law does not provide material conditions laying out the precise balance to be applied in conflict situations but rather necessitates a case-by-case analysis.[65] The CJEU refrains from affirming the primacy of any of fundamental rights in question and therefore does not offer a calculus for evaluating trade-offs between the relevant interests. This has been found inappropriate for it rests on the assumption that all fundamental rights are equal.[66] The CJEU is also criticized for referring issues back to local courts by mere reference to fairness and proportionality. As result, the local courts frequently attempt to determine which of the fundamental rights in question carries most weight and rule accordingly.[67] The crux of the proportionality principle is finding a proper relationship through a balancing of interests without destroying the essence of the right. It is generally accepted that a limitation should not deprive the right of its core elements preventing the exercise of the right. Yet, the court's jurisprudence also falls short in terms of providing clarity regarding the conditions under which the essence of a right is affected. The Court has not substantially identified the essential or core objectives of the rights in question.[68] As such, the current jurisprudence on the principle of proportionality on copyright related matters does not help derive concrete criteria applicable to different cases of conflict but rather provides abstract formulations giving rise to varying interpretations.[69] In addition, the nature of the dispute in the existing copyright rulings is not suitable for comparison with a case of conflict between the transparency demands and IP rights. In the given cases, IP owners, by relying on their exclusive rights, require the other party to act in a certain way. Yet, in case of transparency requirements, leaving aside the direct intervention powers of the DPAs, it is the IP owner who would be required to take action to allow for the exercise of a fundamental right, The specific application of the test depends on the rights in question, the measures at hand and the

---

[62] C-275/06, *Productores de Mu´sica de Espan˜a (Promusicae)* v. *Telefo´nica de Espan˜a SAU* [2008] ECR I-00271; Case C-70/10, *Scarlet Extended SA* v. *Socie´te´ belge des auteurs, compositeurs et e´diteurs SCRL (SABAM)* [2011] ECR I-0000 ;

[63] Peter Teunissen, "The Balance Puzzle: the ECJ's Method of Proportionality Review for Copyright Injunctions" *European Intellectual Property Review,* 2018, 40(9):579-593; Alexander Peukert, "The Fundamental Right to (intellectual) property" in Christophe Geiger (ed.), *Research Handbook on Human Rights and Intellectual Property*, Cheltenham, UK: Edward Elgar, 2015, 132-148, 135; Christina Angelopoulos and Stijn Smet, "Notice-and-Fair-Balance: How to Reach a Compromise between Fundamental Rights in European Intermediary Liability"*Journal of Media Law,* 2016, 8(2): 266-301, 268; Christina Angelopoulos, "Sketching the Outline of a Ghost: The Fair Balance between Copyright and Fundamental Rights in Intermediary Third Party Liability" *Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media* 2015, 17(6): 72-96.

[64] In this regard later judgments *UPC Telekabel*, (ECJ 27 March 2014, C-314/12,) and *McFadden* (ECJ 15 September 2016, C- 484/14) also fail to identify specifics though the latter partially provides useful considerations regarding the effectiveness standard.

[65] Gianclaudio Malgieri. 2016. "Trade Secrets v Personal Data: A possible solution for balancing rights" *International Data Privacy Law,6*(2):102–116.

[66] Peukert (n 63), 135.

[67] Angelopoulos and Smet (n 63).

[68] Orla Lynskey, *The Foundations of EU Data Protection Law*, Oxford, United Kingdom: Oxford University Press, 2015, 161.

[69] EDPS Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the protection of personal data, 19.12.2019, https://edps.europa.eu/sites/edp/files/publication/19-12-19_edps_proportionality_guidelines_en.pdf

relevant circumstances.[70] This entails a case-by-case analysis which makes it difficult to identify generally applicable standards from the European Court's case law.[71]

The proportionality principle presents a more difficult case when trade secrets are at stake. As explained in the *fourth* paper, any detail of the algorithmic process may be treated as a trade secret, and this requires the physical concealment of the relevant information. Given that TS protection ceases to exist once the content of the secret becomes disclosed or known to public, even a simple request to inspect the data collected by an ADM system can be denied based on trade secrets, let alone providing more intricate details of the process. In this respect, even if grounded in one or more of the fundamental rights, transparency claims relating to *access and disclosure*[72] may not escape from infringing the essence of the TS right. This is especially the case where the essence of the right is linked to its market value— an approach likely to tilt the balance in favour of the IP protection. Hence, in case of trade secret claims, the disclosures may definitively remain confined to minor details of ADM systems in use. According to de Laat this renders the transparency provisions of the GDPR (Articles 12,13,15 and 22) toothless.[73] When confronted with trade secrets there seems to be so little margin that balancing could result in favour of transparency requirements.[74] So far, the CJEU has not provided any specific case law addressing a possible conflict between trade secrets protection and Article 22 of the GDPR on contestation. According to Wachter and others, the proportionality test as has been employed by the European Courts would open up some generalities about the system while carefully concealing the concrete details which are vital for effective contestation.[75]. De Laat draws attention to two US cases where trade secret owners' interests prevailed against disclosure demands relating to COMPAS[76] system (a profiling tool used for assessing the chances of recidivism of inmates) and EVAAS (an algorithm for calculating schoolteacher effectiveness for decisions such as bonuses and contract renewal). In both cases, the developers of the systems in question refused to provide any more details beyond what was already known, asserting IP claims both on the source code and the algorithms to keep them secret.[77]

There is growing concern that trade secret protection is transforming from a remedy against competitors into a blanket shield against any type of scrutiny.[78] However, it is yet early to reach a definitive conclusion about how this issue would be resolved before the European courts. Vale and Zanfir-Fortuna report a pending case before the CJEU (referred by the Vienna Regional

---

[70] Considering the relevance to human autonomy and other democratic values, this inquiry extends to the policy question whether algorithms and data used in AI-based systems justify a special IP treatment, namely a lessened protection. Put in other words, the proportionality test could also be seen as a political question in the guise of a legal problem. Gonçalo de Almeida Ribeiro, *The Decline of Private Law: A Philosophical History of Liberal Legalism*, Oxford, UK; Chicago, Illinois: Hart Publishing, 2019, xx.

[71] de Laat (n 60).

[72] 1st layer transparency measures as defined in the *third* paper (Ch.4).

[73] de Laat (n 60), 9-10.

[74] For a more comprehensive evaluation on this matter, *see* Edelson *et al*. (n 51), 36-42.

[75] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation", *International Data Privacy Law,* 2017, 7(2): 76–99, 87 https://doi.org/10.1093/idpl/ipx005.

[76] *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) is a case management and decision support tool developed and owned by Northpointe (now Equivant) used by U.S. courts to assess the likelihood of a defendant becoming a recidivist.

[77] For further details of the COMPAS case, see Monika Zalnieriute, Lyria Bennett Moses, George Williams "The Rule of law and Automation of Government Decision-Making." *The Modern Law Review* 2019, 82 (3): 425–455.

[78] de Laat (n 60), 17.

Administrative Court) which involves the question whether a data controller could invoke trade secret rights to avoid the disclosure of essential information about its credit scoring system and thus prevent the data subject from exercising her/his rights under Article 22 of the GDPR.[79]

Apparently, the question of balancing trade secrets against transparency demands requires a robust and methodological approach to clearly distinguish the elements the disclosure of which would significantly infringe the essence of the TS protection.[80] The judicial and administrative authorities should invest the necessary efforts to ensure that TS claims do not spill over to the elements or information which fall outside the statutory core of the TS protection. In sum, although TS protection may prevent a full disclosure of the algorithm or the source code that was used, for instance in a credit scoring system, it could still be possible to carve out the necessary information such as parameters or input variables together with their effect on the overall score and the reasons why a particular score was assigned. A further solution could be permitting limited disclosure to oversight bodies whose members or employees will be put under a statutory obligation to safeguard the information subject to TS protection.

* * *

An alternative approach (in terms of preventing IP rights from becoming a barrier) may be found in the expanding scholarship which takes the view that current IP rights entrench social divisions and disparities, giving rise to disproportionate protection of IP owners' interests against other legitimate concerns. Up until the emergence of digital technologies there has been a general acceptance of the proprietarian justifications on information and knowledge conferred by IP. Starting from the late 19. century, the notion of IP has expanded from an exceptional 'privilege' to exclude to a 'right' over virtual assets.[81] Today, both the European and the international IP landscape laid down in the TRIPS and other WIPO conventions are predominantly shaped by the tenets of individualism, egalitarianism, and liberalism. However, with the rapid pervasion of data driven technologies, this mostly *utilitarian*[82] regulatory paradigm has been a source of controversy as it gives rise to biased interpretations of IP which dominantly weigh in favour of the commercial interests, giving excessive control over data and information to private parties. It is a widely shared view that commodification of information and

[79] The Austrian Court also asks whether information about the logic involved in automated processing includes the input data, parameters and variables used for profiling; the mathematical formula to calculate the rating; and the enumeration and explanation of each profile category together with an explanation of why the individual was assigned to a particular group/category. Sebastião Barros Vale and Gabriela Zanfir-Fortuna, "Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities" Future of Privacy Forum report, May 2022, 19.

[80] The possible conflict with trade secrets also emerges as a problem under the upcoming EU legislation aiming to foster access and sharing of data. The Data Act proposal which introduces mandatory access rights for data repeatedly refers to trade secrets in various provisions stating that appropriate measures shall be taken to preserve the confidentiality of the trade secrets. However, these references provide almost no guidance about how the provisions of the proposed Act relating to access or sharing of data will be applied to cases where the data in question contains or constitutes a trade secret.

[81] "From the beginning of the development of the international IP regime until towards the end of the 20th century, the concept of intellectual property was about neither 'property' nor 'rights'". Phoebe Li, "Intellectual property for humanity: A manifesto" in Daniel J. Gervais (ed.) *The Future of Intellectual Property*, Cheltenham, UK ; Northampton, Massachusetts: Edward Elgar Publishing Limited, 2021, 9-37, 13. Also see, Rochelle Dreyfuss and Susy Frankel, "From Incentive to Commodity to Asset: How International Law is Reconceptualizing Intellectual Property" *Michigan Journal of International Law,* 2015, 36(4):557; Mark A. Lemley, "Ex Ante Versus Ex Post Justifications for Intellectual Property'", *University Chicago Law Review,* 2004, 71(1):129.

[82] For more on *utilitarian and dignitarian* approaches, see Li *ibid*.

knowledge through expansion of IP (built upon rights-based models and 'rights' narratives) have significantly neglected other underlying values. Hence, a critical approach towards this rights-based regime has emerged—advocating for the enactment of certain legal duties to limit the scope of IP monopolies. [83] As an IP strategy which will optimise the incentives for innovation and public access to information, these scholarly efforts attempt to redefine the concept of "ownership" in light of duties and labilities aiming to mitigate the harmful effects stemming from the exclusive nature of the property rights.[84] Under this approach, the transparency requirements may be formulated as a part of the IP duties— providing a complementary dimension of a sustainable, collaborative, and equitable IP ecosystem. If given legislative or judicial recognition, this could offer a more balanced framework incorporating "duties" to redress the undesirable consequences of exclusive property rights and thus, enable a policy-based and consistent application. As such, the transparency requirements for the purposes of contestation may be identified as *per se* noninfringement cases without the need to apply the proportionality test. Formulating transparency requirements as "IP duties" also aligns with the diginatarian principles (implicit in the foundation of the notion IP) which approach rights as regulatory tools for maximising public interests and social welfare.[85]

Lastly, as an alternative to the current rights-based IP monopolies, "unfair competition doctrine" could offer solutions to overcome difficulties in terms of regulating data and the algorithms within the confines of IP rights or similar regimes based on exclusivity. As a predefined form of tortious conduct, unfair competition rules f particular acts (which harm competitors) and accordingly penalise violation through monetary, administrative or criminal sanctions.[86] Unlike IP rights and trade secrets, unfair competition doctrine does not rest on *erga omnes* exclusivity and does not apply to non-competitors. For instance, under unfair competition rules, individuals may copy and make use of the search results of Google without restriction as long as this does not cause an addressable harm to Google. However, the existence of an exclusive right on the search results as a database (e.g., EU *sui generis* right) would preclude any use of the search results irrespective of any harm to Google. Transparency requirements aiming for contestability and in general legibility of ADM systems, do not relate to activities that compete with the developers or operators of the ADM systems but rather pursue a public or social interest. In search of a legal solution to accommodate transparency demands under the unfair competition doctrine, the question is not whether any access or use of the data or algorithms create an infringement in the abstract. It is rather whether such activities create addressable harms to the current or future commercial exploitation of the materials subject to copyright, *sui generis* right or patent right.

Regarding concerns about the integrity of the systems and the protection of trade secrets, the need for concealment may frequently present a real conflict with competitive interests. This signifies the importance of approaches similar to IP duties explained above which would treat personal data protection, privacy or non-discrimination as prevailing values necessitating a restriction of the property

---

[83] Peter Drahos and Ruth Mayne (eds) *Global Intellectual Property Rights: Knowledge, Access and Development*, Houndmills, Basingstoke, Hampshire; New York: [Oxford, England]: Palgrave Macmillan ; Oxfam, 2002; Keith E. Maskus, *Private Rights and Public Problems: the Global Economics of Intellectual Property in the 21st Century,* Washington, DC: Peterson Institute for International Economics, 2012.
[84] Regarding duties also see Jeremy Waldron, 'Rights in Conflict' *Ethics*, 1989, 99(3): 503–519; Scott Veitch, "The Sense of Obligation" *Jurisprudence,* 2017, 8(3):415–34, 423.
[85] See TRIPS Agreement and the Doha Declaration. Phoebe Li Intellectual property for humanity: A manifesto.
[86] Josef Drexl and others, 'Position Statement of the Max Planck Institute for Innovation and Competition of August 16, 2016 "On the current debate on exclusive rights and access rights to data at the European level"' (Max Planck Institute for Innovation and Competition, 2016) www.ip.mpg.de/en/link/positionpaper-data- 2016-08-16.html

rights. While it is important to ensure that system designers/operators are not forced to disclose more than necessary, it is also vital to prevent any shifting or twisting of commercial interests going so far as to put fundamental rights on equal standing with the necessities of the business model.

The bottom-line is that law should not permit IP rights being used as a pretext or leverage for unsubstantiated secrecy practices. This is particularly the case where data holders bring blanket TS claims aiming to conceal the systems in whole. In this respect, more interdisciplinary research is needed to differentiate between the core aspects of the ADM systems that are of competitive commercial value and other informational elements such as the existence of the system, the purpose for which it was deployed, or the results of an internal impact assessment.[87]

[87] Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker, "Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability", 2018, AInow.

Bibliography

Al-Jazari, Ibn al-Razzaz. 1974 [1204 -1206]. *The Book of Knowledge of Ingenious Mechanical Devices: Kitáb fí ma'rifat al-hiyal al-handasiyya,* (Translated and annotated Donald R. Hill), Dordrecht / Boston: D Reidel Publishing Company.

Allo, Patrick. 2018. "Mathematical Values And The Epistemology Of Data Practices" in Emre Bayamlıoğlu, Irina Baraliuc, Liisa Janssens and Mireille Hildebrandt, (eds), *BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen*, Amsterdam: Amsterdam University Press, 20-23 https://doi.org/10.1515/9789048550180-004

Almada, Marco. 2019. "Human Intervention in Automated Decision-Making: Toward the Construction of Contestable Systems" ICAIL '19: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, 2–11. https:// doi.org/10.1145/3322640.3326699.

Amoore, Louise. 2013. *The Politics of Possibility: Risk and Security beyond Probability*, Durham: Duke University Press.

Amoore, Louise and Volha Piotukh. 2015."Life Beyond Big Data: Governing with Little Analytics", *Economy and Society*, 44(3): 341-366.

Ananny, Mike and Kate Crawford,. 2018. "Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability", *New Media & Society*, 20(3):973-989.

Anderson, Ben. 2010. "Preemption, precaution, preparedness: Anticipatory action and future geographies" *Progress in Human Geography* 34(6): 777-798.

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", *Wired Magazine* (23 June 2008) http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory.

Angelopoulos, Christina. 2015. "Sketching the Outline of a Ghost: The Fair Balance between Copyright and Fundamental Rights in Intermediary Third Party Liability" *Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media*, 17(6): 72-96.

Angelopoulos, Cristina. 2016. "Database Directive" in Thomas Dreier and PB Hugenholtz (eds), *Concise European copyright law,* (2nd ed) Alphen aan den Rijn, The Netherlands: Kluwer Law International.

Angelopoulos, Christina, and Stijn Smet. 2016. "Notice-and-Fair-Balance: How to Reach a Compromise between Fundamental Rights in European Intermediary Liability"*Journal of Media Law* 8(2): 266-301.

Aplin, Tanya. 2015. "Right to Property and Trade Secrets" in Christophe Geiger (ed), *Research Handbook on Human Rights and Intellectual Property,* Cheltenham, UK: Edward Elgar Publishing. 421-437.

Aplin, Tanya. 2017. "Trading Data in the Digital Economy: Trade Secrets Perspective" in Sebastian Lohsse, Reiner Schulze and Dirk Staudenmayer (eds), *Trading Data in the Digital Economy: Legal Concepts and Tools (*1st ed.) Baden-Baden, Germany : [Oxford, England]: Nomos ; Hart Publishing, 59-72.

Argenton, C. and J Prüfer. 2012. "Search Engine Competition with Network Externalities", *Journal of Competition Law and Economics* 8(1): 73–105. https://doi.org/10.1093/joclec/nhr018.

Ashley, Kevin. 2017. *Artificial Intelligence and Legal Analytics – New Tools for Law Practice in the Digital Age*, Cambridge: Cambridge University Press.

Asscher, Lodewijk. 2006 '"Code" as Law. Using Fuller to Assess Code Rules' in Egbert Dommering and Lodewijk Asscher (eds), Coding Regulation – Essays on the Normative Role of Information Technology The Hague : TMC Asser, 61–90.

Baer, Tobias. 2019. *Understand, Manage, and Prevent Algorithmic Bias: A Guide for Business Users and Data Scientists*. New York: Apress.

Ballardini, Rosa Maria. 2010. "Scope of IP Protection for the Functional Elements of Software' *In Search of New IP Regimes,* IPR University Center 2010, 27-62.

Bambauer Jane and Tal Zarsky T. 2018. "The Algorithm Game", *Notre Dame Law Review 94*(1), 1–48.

Baker, Stephen. 2009. *The Numerati*, Boston: Mariner Books.

Band, Jonathan. 2018. "The Global API Copyright Conflict" 31 *Harvard Journal of Law & Technology,* Special Issue Spring 2018, 615-636.

Bannister, Frank and Regina Connolly. 2020. "Administration by algorithm: a risk management framework" *Information Polity*, 25(4):471-490.

Banterle, Francesco. 2018. "The Interface between Data Protection and IP Law. The Case of Trade Secrets and the Database *sui generis* Right in Marketing Operations, and the Ownership of Raw Data in Big Data Analysis" in Bakhoum, Mor, Beatriz Conde Gallego, Mark-Oliver Mackenrodt, and Gintarė Surblytė-Namavičienė (eds), *Personal Data in Competition, Consumer Protection and Intellectual Property Law: Towards a Holistic Approach?* Berlin, Heidelberg: Springer Berlin Heidelberg, 411–443. https://doi.org/10.1007/978-3-662-57646-5.

Banterle, Francesco. 2020. "Data Ownership in the Data Economy: A European Dilemma" in Synodinu, Tatianē-Elenē, Philippe Jougleux, Christiana Markou, and Thalia Prastitou, eds. *EU Internet Law in the Digital Era: Regulation and Enforcement*. Cham, Switzerland: Springer, 199-225.

Bamberger, Kenneth A. 2006. "Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State", Vol.56 *Duke Law Journal* No.2:377-468.

Barocas, Solon and Andrew Selbst, 2016. "Big Data's Disparate Impact" *California Law Review*, Vol. 104, http://ssrn.com/abstract=2477899

Baskarada, Sasa, and Andy Koronios. 2013. "Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and Its Quality Dimension", *Australasian Journal of Information Systems* 18, no. 1 https://doi.org/10.3127/ajis.v18i1.748.

Bateson, Gregory. 1987. *Steps to an Ecology of Mind,* (1972 1st ed.) Jason Aronson Inc, Northvale, NJ, and London.

Bauman, Zygmunt, Didier Bigo, Paulo Esteves, Elspeth Guild, Vivienne Jabri, David Lyon, and R. B. J. Walker. 2014."After Snowden: Rethinking the Impact of Surveillance". *International Political Sociology* 8, no. 2 (June 2014): 121–44.

Bayamlıoğlu, Emre and Ronald Leenes, 2018. "The 'rule of law' implications of data-driven decision- making: A techno-regulatory perspective", *Law, Innovation and Technology* 10, no. 2: 295–313. https://doi.org/10.1080/17579961.2018.1527475.

Bayamlıoğlu, Emre. 2018. "Contesting Automated Decisions", *European Data Protection Law Review* 4: 433–446.

Bayamlıoğlu, Emre. 2021. "The Right to Contest Automated Decisions under the General Data Protection Regulation : Beyond the So-called 'Right to Explanation'"*Regulation & Governance* 16(4): 1058–78. https://doi.org/10.1111/rego.12391.

Beckerman-Rodau, Andrew. 2002. "The Choice Between Patent Protection and Trade Secret Protection: A Legal and Business Decision", 84 *Journal of the Patent & Trademark Office Society*, 371.

Beer, David. 2015. "Productive measures: Culture and measurement in the context of everyday neoliberalism" *Big Data & Society*, January–June:1–12.

Bently, Lionel, Brad Sherman, Denis Borges Barbosa, Shamnad Basheer, Coenraad Visser and Richard Gold, 2010. "Study on Exclusions from Patentability and Exceptions and Limitations to Patentees' Rights" WIPO Standing Committee on the Law of Patents, SCP/15/3 Annex I, 2010)

Bently, Lionel, and Brad Sherman. 2014. *Intellectual Property Law*, (4th edition) Oxford, United Kingdom: Oxford University Press.

Berry, David M. 2011. *The Philosophy of Software: Code and Mediation in the Digital Age*. Basingstoke, Hampshire, New York: Palgrave Macmillan.

Bertea, Stefano. 2009. *The Normative Claim of Law Law and Practical Reason*, Oxford; Portland, Or: Hart Publishing.

Bibri, Simon Elias. 2015. *The Human Face of Ambient Intelligence: Cognitive, Emotional, Affective, Behavioral And Conversational Aspects,* Paris: Atlantis Press https://doi.org/10.2991/978-94-6239-130-7.

Binns, Ruben. 2017. "Algorithmic Accountability and Public Reason" *Philosophy & Technology*:1-14. doi: 10.1007/s13347-017-0263-5.

Black, Julia. 2002. "Critical Reflections on Regulation" 27 *Australian Journal of Legal Philosophy* 1.

Bollier, David. 2010. *The Promise and Peril of Big Data,* Washington: The Aspen Institute.

Bone, Robert G. 2014. "The (Still) Shaky Foundations of Trade Secret Law", 92 *Texas Law Review,* 1803.

Bone, Robert G. 2011."Trade Secrecy, Innovation and the Requirement of Reasonable Secrecy Precautions" in Rochelle C. Dreyfuss and Katherine J. Strandburg (eds) *The Law and Theory of Trade Secrecy*, Cheltenham, UK; Northampton, MA: Edward Elgar.

Booi, Mairan (ed). 2011. *The Panoptic Writings*, London: Verso.

Borghi, Maurizio and Stavroula Karapapa. 2015. "Contractual restrictions on lawful use of information: sole-source databases protected by the back door?" *European Intellectual Property Review*, 37(8):505-514.

Braitenberg, Valentinoi. 1977. *On the Texture of Brains: An Introduction to Neuroanatomy for the Cybernetically Minded,* Berlin Heidelberg: Springer-Verlag.

Brighenti, Andrea. 2007. "Visibility: A Category for the Social Sciences", *Current Sociology*, 55(3): 323–342. https://doi.org/10.1177/0011392107076079.

Brown, Ian and Chris Marsden. 2013. *Regulating Code. Good Governance and Better Regulation in the Information Age,* Cambridge, MA, London: MIT Press.

Brownsword, Roger and Morag Goodwin. 2012. *Law and the Technologies of the Twenty-First Century. Text and Materials, Cambridge, UK ; New York:* Cambridge University Press.

Brownsword, Roger. 2018. "So What Does the World Need Now? Reflections on Regulating Technologies" in R Brownsword and K Yeung (eds), *Regulating Technologies Legal: Futures, Regulatory Frames and Technological Fixes*. Oxford, UK: Hart Publishing, 23–48.

Brownsword, Roger. 2005. "Code, Control, and Choice: Why East is East and West is West" *Legal Studies* 25, no. 1:1–21. https://doi.org/10.1111/j.1748-121X.2005.tb00268.x

Bryson, Joanna. 2020. "The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation" in Markus D. Dubber, Frank Pasquale, Sunit Das (eds), *The Oxford Handbook Of Ethics Of AI*, New York, NY: Oxford University Press, 1-24.

Bucher, Taina. 2018. *If...Then: Algorithmic Power and Politics*. New York: Oxford University Press.

Bunge, Mario. 2012. *Evaluating Philosophies,* Dordrecht: Springer Netherlands.

Burk, Dan L..2021. "Algorithmic Legal Metrics", *Notre Dame Law Review* 96(3):1147-1201.

Burrell, Jenna. 2016. "How the machine 'thinks': Understanding opacity in machine learning algorithms", *Big Data & Society*, January–June: 1–12.

Burri, Thomas. 2017. "Machine Learning and the Law: 5 Theses" *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2927625.

Burri, Mira and Ingo Meitinger. 2014. "The Protection of Undisclosed Information: Commentary of Article 39 TRIPS" in Thomas Cottier and Pierre Véron (eds), *Concise International and European IP Law: TRIPS, Paris Convention, European Enforcement and Transfer of Technology,* The Hague: Kluwer Law International.

Buscema, Massimo and William J. Tastle (eds). 2013. *Intelligent Data Mining in Law Enforcement Analytics: New Neural Networks Applied to Real Problems*. New York: Springer.

Bygrave, Lee. 2001. "Automated Profiling: Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling", *Computer Law & Security Review* 17(1): 17–24.

Calo, M.Ryan. 2013. "Digital market manipulation" *George Wash Law Rev* 82:995.

Calude, Cristian S. and Giuseppe Longo. 2017. "The Deluge of Spurious Correlations in Big Data", *Foundations of Science* 22(3): 595–612.

Castañeda, Hector-Neri. 1970. "On the Semantics of the Ought-to-Do", *Synthese* 21(3–4):449–468. https://doi.org/10.1007/BF00484811.

Cath, Corinne. 2018. "Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2133: 20180080. https://doi.org/10.1098/rsta.2018.0080.

Chandler, David. 2015. "A World without Causation: Big Data and the Coming of Age of Posthumanism", *Millennium: Journal of International Studies* 43(3): 833–851.

Chandler, David. 2018. *Ontopolitics in the Anthropocene: An Introduction to Mapping, Sensing and Hacking,* Abingdon, Oxon; New York, NY:Routledge, 2018.

Chandler, David. 2019. "Digital Governance in the Anthropocene: The Rise of the Correlational Machine" in David Chandler and Christian Fuchs (eds.) *Digital Objects, Digital Subjects*, *Interdisciplinary Perspectives on Capitalism, Labour and Politics in the Age of Big Data*, London: University of Westminster Press, 23–42.

Celeste, Edoardo and Giovanni De Gregorio. 2022. "Digital Humanism: The Constitutional Message of the GDPR", *Global Privacy Law Review*, Issue 1, 4-18.

Chamayou, Grégoire. 2015. *A Theory of the Drone* (trans. Janet Lloyd). The New Press.

Cheney-Lippold, John, 2017. *We Are Data: Algorithms and the Making of Our Digital Selves*, New York: NYU Press.

Citron, Danielle Keats. 2007. "Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age", *Southern California Law Review,* 80(2): 241-297.

Citron, Danielle Keats. 2008. "Technological Due Process", *Washington University Law Review* 85(6):1249-1313. https://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2

Citron, Danielle Keats and Frank Pasquale. 2014. "The scored society", *Washington Law Review* 89(1):1–33

Cofone, Ignacio, and Katherine J. Strandburg. 2019. "Strategic Games and Algorithmic Secrecy", *McGill Law .Journal,* 64.4, 623.

Coglianese, Cary and David Lehr. 2019. "Transparency and Algorithmic Governance", Faculty Scholarship at Penn Law 2123.https://scholarship.law.upenn.edu/faculty_scholarship/2123

Colakides, Yiannis, Marc Garrett, Inte Gloerich (eds). 2019. *State Machines: Reflections and Actions at the Edge of Digital Citizenship, Finance, and Art,* Amsterdam: Institute of Network Cultures.

Colonna, Liane. 2013."A Taxonomy and Classification of Data Mining" 16 *SMU Science and Technology Law Review*, 309.

Comandé, Giovanni. 2017. "Regulating Algorithms' Regulation? First Ethico-Legal Principles, Problems and Opportunities of Algorithms" in Tania Cerquitelli, Daniele Quercia and Frank Pasquale (eds), *Transparent Data Mining for Small and Big Data,* New York, NY: Springer Science+Business Media, 169-206.

Compagnucci, Marcelo Corrales. *Big Data, Databases and 'Ownership' Rights in the Cloud*, Singapore: Springer, 2020.

Cubitt, Sean. 2017. *Finite Media: Environmental Implications of Digital Technologies*, Durham: Duke University Press.

Cupitt, Philip. 2019. "Patenting Artificial Intelligence at the European Patent Office" (*Marks&Clerk,* 11 April 2019) <*https://www.marks-clerk.com/Home/Knowledge-News/Articles/Patenting-Artificial-Intelligence-at-the- European.aspx#.YDwfAS2cY1I*>.

Dancy, Jonathan. 2006. "Non-naturalism" in David Copp (ed), *The Oxford Handbook of Ethical Theory,* New York: Oxford University Press, 122–145.

Danks, David and Alex John London. 2017. "Algorithmic Bias in Autonomous Systems" in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 4691–97. Melbourne, Australia: https://doi.org/10.24963/ijcai.2017/654.

Davenport, Thomas H. 2018. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. Cambridge, MA: MIT Press.

Davison, Mark J. 2003. *The Legal Protection of Databases*. Cambridge: Cambridge University Press.

De Hert, Paul, Vagelis Papakonstantinou, Gianclaudio Malgieri, Laurent Beslay, and Ignacio Sanchez. 2018. "The Right to Data Portability in the GDPR: Towards User-Centric Interoperability of Digital Services'. *Computer Law & Security Review* 34(2): 193–203. https://doi.org/10.1016/j.clsr.2017.10.003.

Delacroix, Sylvie. 2006. *Legal Norms and Normativity: An Essay in Genealogy*, Oxford; Portland, OR: Hart Publishing.

Deltorn, Jean-Marc. 2017. "Deep Creations: Intellectual Property and the Automata" 4(3) Front. Digit. Humanit.

Derclaye, Estelle, 2021. "The Database Directive" in Irini Stamatoudi and Paul Torremans (eds), *EU Copyright Law: A Commentary,* Cheltenham, UK ; Northampton, MA: Edward Elgar Publishing, 216–254.

De Ridder Jeroen. 2006. "The Inherent Normativity of Technological Explanations" Techné: Research in Philosophy and Technology 10(1): 79–94.

Desai, Deven and Joshua Kroll. 2017. "Trust but Verify: A Guide to Algorithms and the Law", *Harvard Journal of Law & Technology* 31: 2–64.

Desrosières, Alain. 1998. *The Politics of Large Numbers: A History of Statistical Reasoning*, Cambridge, Mass: Harvard University Press.

Diakopoulos, Nicholas. 2014. "Algorithmic Accountability: Reporting On The Investigation of Black Boxes" Tow Center for Digital Journalism, Columbia Journalism School: https://academiccommons.columbia.edu/doi/ 10.7916/D8ZK5TW2.

Diakopoulos, Nicholas. 2020. "Accountability, Transparency, and Algorithms", in Markus D. Dubber, Frank Pasquale, Sunit Das (eds), *The Oxford Handbook of Ethics Of AI*, New York, NY: Oxford University Press, 197-213.

Dormehl, Luke. 2015. *The Formula: How Algorithms Solve All Our Problems and Create More*, New York: Penguin Group (USA) LLC.

Doshi-Velez, Finale, Mason Kortz, Ryan Budish, Christopher Bavitz, Samuel J. Gershman, David O'Brien, Stuart Shieber, Jim Waldo, David Weinberger, and Alexandra Wood. 2017. "Accountability of AI Under the Law: The Role of Explanation", Berkman Klein Center for Internet & Society working paper, *SSRN Electronic Journal*, 2017. https://doi.org/10.2139/ssrn.3064761.

Drahos, Peter, Ruth Mayne and Oxfam GB (eds). 2002. *Global Intellectual Property Rights: Knowledge, Access, and Development*. Houndmills, Basingstoke, Hampshire ; New York : [Oxford, England]: Palgrave Macmillan; Oxfam.

Drexl, Josef. 2018. *Data Access and Control in the Era of Connected Devices,* Study on Behalf of the European Consumer Organisation BEUC.

Drexl, Josef. 2017. "Designing Competitive Markets for Industrial Data – Between Propertisation and Access, 8 *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 257 para 1. https://www.jipitec.eu/issues/jipitec-8-4-2017/4636/JIPITEC_8_4_2017_257_Drexl

Drexl, Josef, Reto M. Hilty, Luc Desaunettes, Franziska Greiner, Daria Kim, Heiko Richter, Gintarė Surblytė and Klaus Wiedemann. 2016. "Data Ownership and Access to Data -Position Statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the Current European Debate" Max Planck Institute for Innovation & Competition Research Paper No. 10 https://ssrn.com/abstract=2833165

Dreyfus, Hubert. 1965. *Alchemy and Artificial Intelligence,* Rand Corporation.

Dreyfuss, Rochelle and Susy Frankel. 2015. "From Incentive to Commodity to Asset: How International Law is Reconceptualizing Intellectual Property" *Michigan Journal of International Law* 36(4), 557.

Ducuing, Charlotte, Thomas Margoni and Luca Schirru (eds). 2022. *White Paper on the Data Act Proposal*, CiTiP Working Paper Series 26.02.22.

Economides, Constantin P**.** 2010 "Content of the Obligation: Obligations of Means and Obligations of Result" (Ch.26) in James Crawford, Alain Pellet, Simon Olleson, Kate Parlett (eds), *The Law of International Responsibility*, New York: Oxford University Press.

Edelson, Laura, Inge Graef, and Filippo Lancieri. 2023. "Access to Data and Algorithms: For an Effective DMA and DSA Implementation" (CERRE, March 2023), https://cerre.eu/publications/access-to-data-and-algorithms-for-an-effective-dma-and-dsaimplementation

Edwards, Lilian and Michael Veale. 2017. "Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For" *Duke Law and Technology Review*, 16(1):1-65.

Edwards, Lilian. 2020. "The EU AI Act proposal: a summary of its significance and scope" Ada Lovelace Institute,. https://www.adalovelaceinstitute.org/ resource/eu-ai-act-explainer/

Eliband, Malin, Hanna Schneider and Daniel Buschek. 2018. "Normative vs Pragmatic: Two Perspectives on the Design of Explanations in Intelligent Systems" ExSS '18, Tokyo, Japan, 11 March 2018.

Erwig, Martin. 2017. *Once upon an Algorithm: What Stories Can Teach Us about Computation,* MIT Press.

Eslami, Motahhare, Amirhossein Aleyasen, Roshanak Zilouchian Moghaddam, and Karrie Karahalios. 2014. "Friend Grouping Algorithms for Online Social Networks: Preference, Bias, and Implications" in LM Aiello and D McFarland (eds)*, Social Informatics: 6th International Conference, Socinfo 2014,* Barcelona, Spain.. *Lecture Notes in Computer Science,* New York: Springer*,* 34-49.

Eyert, Florian, Florian Irgmaier and Lena Ulbricht. 2022. "Extending the Framework of Algorithmic Regulation. The Uber Case", *Regulation & Governance* 16(1): 23–44. https://doi.org/10.1111/rego.12371.

Feenberg, Andrew. 2009. "Critical Theory of Technology" in Friis, Jan Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendricks, (eds), *A Companion to the Philosophy of Technology*, Chichester, UK ; Malden, MA: Wiley-Blackwell, 146-154.

Feigenbaum, E.A.1977. "The Art of Artificial Intelligence: I Themes and Case Studies of Knowledge Engineering. Technical Report" (UMI Order Number: CS-TR-77–621, Stanford University.

Felten, Edward. 2017. "What does it mean to ask for an 'explainable' algorithm?" (*Freedom to Tinker*, https://freedom-to -tinker.com/2017/05/31/what-does-it-mean-to-ask-for-an -explainable-algorithm/

Felzmann, Heike, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2019. "Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns" *Big Data & Society* 6(1): 1-14.

Fenster, Mark.2015. "Transparency in Search of a Theory", *European Journal of Social Theory* 18, no. 2:150–67. https://doi.org/10.1177/1368431014555257

Ferraris, Valeria and Bosco, Francesca and Cafiero, G. and D'Angelo, Elena and Suloyeva, Y. 2013. "Defining Profiling". https://ssrn.com/abstract=2366564 or http://dx.doi.org/10.2139/ssrn.2366564

Ferretti, Federico. 2013. "The Legal Framework of Consumer Credit Bureaus and Credit Scoring in the European Union: Pitfalls and Challenges - Overindebtedness, Responsible Lending, Market Integration, and Fundamental Rights." *Suffolk University Law Review* 46: 791.

Flach, Peter A. 2012. *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge; New York: Cambridge University Pres.

Fleming, James E., (ed). 2011. *Getting to the Rule of Law*. Nomos: Yearbook of the American Society for Political and Legal Philosophy 50. New York: New York University Press.

Franssen, Maarten. 2014. 'The Good, the Bad, the Ugly... and the Poor: Instrumental and Non-instrumental Value of Artefacts*'* in P Kroes and P Verbeek (eds), *The Moral Status of Technical Artefacts*, Dordrecht: Springer.

Franssen, Maarten. 2009. "Artefacts and normativity" in Anthonie Meijers (ed.), *Philosophy of Technology and Engineering Sciences,* 1. ed. Handbook of the Philosophy of Science 9. Amsterdam Heidelberg: Elsevier, 923-952.

Friedler, Sorelle, A. Carlos Scheiddegger and Suresh Venkatasubramanian. 2016. 'On the (im)possibility of fairness' arX- iv:1609.07236v1. https://doi.org/10.48550/arXiv.1609.07236

Friedman, Batya, and Helen Nissenbaum. 1996. "Bias in Computer Systems", *ACM Transactions on Information Systems* 14, no. 3:330–47. https://doi.org/10.1145/230538.230561.

Friedman, David D, William M Landes, and Richard A Posner. 1991. "Some Economics of Trade Secret Law" *Journal of Economic Perspectives* 5 (1): 61–72. https://doi.org/10.1257/jep.5.1.61.

Fromer, Jeanne C. 2019. "Machines as the New Oompa-Loompas: Trade Secrecy, the Cloud, Machine Learning, and Automation" 94 *New York University Law Review*, 706.

Fuchs, Christian. 2019. "Beyond Big Data Capitalism, Towards Dialectical Digital Modernity: Reflections on David Chandler's Chapter" in David Chandler and Christian Fuchs (eds.), *Digital Objects, Digital Subjects: Interdisciplinary Perspectives on Capitalism, Labour and Politics in the Age of Big Data*. London: University of Westminster Press, 43–51.

Gonzalez Fuster, Gloria. 2018. "Transparency as translation in data protection" in Emre Bayamlıoğlu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt, (eds),*BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen*, Amsterdam: Amsterdam University Press, 52-55.

Gonzalez Fuster, Gloria. 2014. *The Emergence of Personal Data Protection as a Fundamental Right of the EU*, Cham: Springer International Publishing.

Gallie, Walter Bryce. 1955. "Essentially Contested Concepts," in *Proceedings of the Aristotelian Society* 56: 167–198.

Geiger, Christophe and Giancarlo Frosio, 2018. "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects" CEIPI Research Paper 2018-02 < *http://ssrn.com/abstract_id=3160586*

George, Frank. 1977. *Machine Takeover, The Growing Threat to Human Freedom in a Computer Controlled Society*, Elmsford, N.Y: Pergamon Press.

Gervais, Daniel. 2009. "The Tangled Web of UGC: Making Copyright Sense of User-Generated Content" *Vanderbilt Journal of Entertainment and Technology Law* 11(4): 841-870.

Gervais, Daniel. 2019. "Exploring the Interfaces Between Big Data and Intellectual Property Law" *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 10(3): para 1.

Gervais, Daniel. 2021. "Introduction to the future of intellectual property" in Daniel Gervais (ed,) *The Future of Intellectual Property*, Cheltenham, UK ; Northampton, Massachusetts: Edward Elgar Publishing Limited,1-7.

Gitelman, Lisa (ed.). 2013. *"Raw Data" Is an Oxymoron*, Cambridge, Massachusetts: MIT Press.

Giudice, Michael. 2015. *Understanding the Nature of Law. A Case for Constructive Conceptual Explanation,* Cheltenham, UK: Edward Elgar Publishing.

Goffey, Andrew. 2008. "Algorithm" in Matthew Fuller (ed.), *Software Studies A Lexicon,* Cambridge, Massachusetts, London, England: MIT Press, 15-36.

Gollnick, Clare. 2018. "Induction Is Not Robust to Search" in Emre Bayamlıoğlu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt, (eds),*BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen*, Amsterdam: Amsterdam University Press,106–111.

Graef, Inge. 2015. "Market Definition and Market Power in Data: The Case of Online Platforms", *World Competition* 38(4) :473–505. https://doi.org/10.54648/WOCO2015040.

Grimmelmann, James. 2005. "Regulation by Software", 114 *Yale Law Journal,* 1719.

Grimmelmann, James. 2012. "Three Theories of Copyright in Ratings," 14 *Vanderbilt Journal of Entertainment and Technology Law* 851.

Guibault, Lucie M.C.R. 2010. Why Cherry-Picking Never Leads to Harmonisation: The Case of the Limitations on Copyright under Directive 2001/29/EC" 1 *Journal of Intellectual Property, Information Technology and E-Commerce Law,* 55, para 1.

Guibault, Lucie M.C.R. 2002. *Copyright Limitations and Contracts: An Analysis of the Contractual Overridability of Limitations on Copyright*, The Hague; Boston: Kluwer Law International.

Guimerà, Roger, and Marta Sales-Pardo. 2011. "Justice Blocks and Predictability of U.S. Supreme Court Votes". Yamir Moreno (ed.). *PLoS ONE* 6, no. 11: e27188.

Guinchard, Audrey. 2017. "Contextual Integrity and EU Data Protection Law: Towards a More Informed and Transparent Analysis" *SSRN Electronic Journal*, 2017. https://doi.org/10.2139/ssrn.2946772.

Gunning, David. 2018. "Explainable Artificial Intelligence (XAI)" (Defence Advanced Research Projects Agency, DARPA/I2O) <https:// www.darpa.mil/program/explainable-artificial-intelligence>.

Gutwirth, Serge, Ronald Leenes, and Paul de Hert (eds). 2015. *Reforming European Data Protection Law*. Vol. 20. Law, Governance and Technology Series. Dordrecht: Springer Netherlands, https://doi.org/10.1007/978-94-017-9385-8.

Hand, David J. 2006. "Classifier Technology and the Illusion of Progress'. *Statistical Science*21, no.1.: 1–14. https://doi.org/10.1214/088342306000000060.

Hansen, Hans Krause, and Mikkel Flyverbom. 2015. "The Politics of Transparency and the Calibration of Knowledge in the Digital Age", *Organization* 22, no. 6: 872–89. https://doi.org/10.1177/1350508414522315.

Harizanov, Valentina S., Norma B. Goethe and Michèle Friend. 2007. "Introduction to The Philosophy and Mathematics of Algorithmic Learning Theory" in Michèle Friend, Norma B. Goethe and Valentina .S. Harizanov (eds.), *Induction, Algorithmic Learning Theory, and Philosophy*, Dordrecht: Springer, 1-24.

Hawkes, Terence. 2003. *Structuralism and Semiotics*. 2. ed., Reprinted. New Accents. London: Routledge.

Heald, David. 2006. "Varieties of Transparency" in Christopher Hood and David Heald (eds.). *Transparency the Key to Better Governance,* Oxford: Oxford University Press, 24-43.

Henshaw, John M. 2006. *Does Measurement Measure Up? How Numbers Reveal and Conceal the Truth*, Baltimore: Johns Hopkins University Press.

Herian, Robert. 2022. *Data: New Trajectories in Law*. New Trajectories in Law. Milton Park, Abingdon, Oxon ; New York, NY: Routledge.

Higgins, Brian. 2019. "The Role of Explainable Artificial Intelligence in Patent Law" *Intellectual Property & Technology Law Journal*, 31(3).

Hildebrandt, Mireille. 2008. "Defining Profiling: A New Type of Knowledge?" in Mireille Hildebrandt and Serge Gutwirth (eds.) *Profiling the European Citizen*, Dordrecht: Springer Netherlands, 17–45.

Hildebrandt, Mireille. 2008. "A Vision of Ambient Law" in Roger Brownsword and Karen Yeung (eds), *Regulating Technologies*, Oxford; Portland, Or: Hart, 175–192.

Hildebrandt, Mireille. 2009. 'Technology and the end of law' in E Claes, W Devroe and B Keirsbilck (eds), *Facing The Limits of The Law,* Dordrecht: Springer, 443-464.

Hildebrandt, Mireille. 2010. "Law at a Crossroads: Losing the Thread or Regaining Control? The Collapse of Distance in Real Time Computing" in Morag Goodwin, Bert-Jaap Koops and Ronald Leenes (eds), *Dimensions of Tech- nology Regulation*, Nijmegen: Wolf Legal Publishers.

Hildebrandt, Mireille and Bert-Jaap Koops. 2010. "The Challenges of Ambient Law and Legal Protection in the Profiling Era" 73 *Modern Law Review* LR 428.

Hildebrandt, Mireille. 2011. "A multifocal view of human agency in the era of autonomic computing" in Mireille Hildebrandt and Antoinette Rouvroy (eds), *Law, Human Agency, and Autonomic Computing: The Philosophy of Law Meets the Philosophy of Technology, Milton Park, Abingdon: Routledge,* 1-11.

Hildebrandt, Mireille, 2013. "Profile transparency by design? Re-enabling double contingency" in M Hildebrandt and K de Vries (eds), *Privacy, Due Process and the Computational Turn*, Abingdon, Oxon, [England] ; New York Routledge, 221-246.

Hildebrandt, Mireille, and Katja de Vries (eds). 2013. *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Abingdon, Oxon, [England] ; New York: Routledge.

Hildebrandt, Mireille, and Laura Tielemans. 2013. "Data Protection by Design and Technology Neutral Law", *Computer Law & Security Review* 29, no. 5: 509–21. https://doi.org/10.1016/j.clsr.2013.07.004.

Hildebrandt, Mireille. 2015. "Radbruch's Rechtsstaat and Schmitt's Legal Order: Legalism, Legality, and the Institution of Law", *Critical Analysis of Law* 2:1.

Hildebrandt, Mireille. 2015. *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Cheltenham, UK: Edward Elgar Publishing.

Hildebrandt, Mireille. 2016. 'Law as Information in the Era of Data-Driven Agency'. *The Modern Law Review* 79 (1): 1–30. doi:10.1111/1468-2230.12165.

Hildebrandt, Mireille. 2016. 'The New Imbroglio. Living with Machine Algorithms' in Liisa Janssens (ed), *The Art of Ethics in the Information Society. Mind you.* Amsterdam: Amsterdam University Press, 55– 60. DOI: https://doi.org/10.25969/mediarep/13395.

Hildebrandt, Mireille. 2018. "Law As Computation in the Era of Artificial Legal Intelligence. Speaking Law to the Power of Statistics" *University of Toronto Law Journal* Vol. 68 Issue supplement 1, January 2018, 12-35.

Hildebrandt, Mireille. 2018. "Preregistration of machine learning research design. Against P-hacking" in Emre Bayamlıoğlu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt, (eds),*BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen*, Amsterdam: Amsterdam University Press,102-105.

Hildebrandt, Mireille. 2019. "Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning" *Theoretical Inquiries in Law*, vol. 20, no. 1: 83-121. https://doi.org/10.1515/til-2019-0004

Holland, John H. 1995. *Hidden Order: How Adaptation Builds Complexity*. Helix Books. Reading, Mass: Addison-Wesley.

Hong, Sun-ha. 2020. *Technologies of Speculation: The Limits of Knowledge in a Data-Driven Society*. New York: New York University Press.

Hood, Christopher. 1983. *The Tools of Government,* London: Macmillan.

Hood, Christopher, Henry Rothstein, and Robert Baldwin. 2001. *The Government of Risk: Understanding Risk Regulation Regimes*. Oxford; New York: Oxford University Press.

Hood, Christopher. 2006. "Transparency in Historical Perspective" in Christopher Hood and David Heald (eds), *Transparency: The Key to Better Government?* Oxford ; New York: Oxford University Press, 2-23.

Hoepman, Jaap-Henk. 2018. "Transparency as Translation in Data Protection" in Emre Bayamlıoğlu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt, (eds),*BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen*, Amsterdam: Amsterdam University Press, 102–105.

Hoffmann-Riem, Wolfgang. 2020. "Artificial Intelligence as a Challenge for Law and Regulation" in Thomas Wischmeyer, and Timo Rademacher (eds.), *Regulating Artificial Intelligence*, Cham, Switzerland: Springer, 1–32.

Horkheimer, Max. 2004. *Eclipse of Reason*. Rev. ed. London; New York: Continuum.

Houkes, W. N. 2013. "Rules, Plans and the Normativity of Technological Knowledge" in Marc J. de Vries, Sven Ove Hansson, and Anthonie Meijers, (eds), *Norms in Technology*, Philosophy of Engineering and Technology, v. 9. Dordrecht; New York: Springer, 35-54.

Hugenholtz, P. Bernt. 1998. "Implementing the European Database Directive" in Jan J.C. Kabel and Gerard J.H.M. Mom (eds), *Intellectual Property and Information Law, Essays In Honour Of Herman Cohen Jehoram,* The Hague; Boston: Kluwer Law International, 183-200.

Hugenholtz, P. Bernt. 2016. "Something Completely Different: Europe's Sui Generis Database Right" in Susy Frankel and Daniel Gervais (eds), *The Internet and the Emerging Importance of New Forms of Intellectual Property,* the Netherlands*:* Kluwer Law International, 205-222.

Hugenholtz, P. Bernt. 2018. "Against 'Data Property"in Hanns Ullrich, Peter Drahos and Gustavo Ghidini (eds), *Kritika: Essays on Intellectual Property,* vol. 3: 48-71*,* Cheltenham, UK: Edward Elgar Publishing.

Husovec, Martin. 2019. "The Essence of Intellectual Property Rights Under Article 17(2) of the EU Charter", *German Law Journal, 20*(6), 840-863. doi:10.1017/glj.2019.65

Ihde, Don. 1990. *Technology and the Lifeworld. From Garden to Earth*, Bloomington: Indiana University Press.

Iliadis, Andrew and Federica Russo. 2016. "Critical Data Studies: An Introduction'. *Big Data & Society* 3, no.2.

Introna, Lucas and Niall Hayes. 2011. "On Sociomaterial Imbrications: What plagiarism detection systems reveal and why it matters" 21 *Information and Organisation,* 107-122.

Introna, Lucas. 2016. "Algorithms, Governance, and Governmentality: On Governing Academic Writing." *Science, Technology, & Human Values* 41(1): 17–49. http://www.jstor.org/stable/43671281.

Ioannidis, Stavros and Stathis Psillos. 2013. "Mechanisms, Counterfactuals, and Laws" in Stuart Glennan and Phyllis Illari (eds), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, Ch.11, London ; New York: Routledge.

Jacobs, Adam. 2009. "The Pathologies of Big Data", *Communications of the ACM,* 52: 36-44.

Jasanoff, Sheila, and Hilton R Simmet. 2017. "No Funeral Bells: Public Reason in a "Post-Truth" Age". *Social Studies of Science* 47, no. 5: 751–70. https://doi.org/10.1177/0306312717731936.

Jin, Hyunjong Ryan. 2018. "Think Big! The Need For Patent Rights In The Era Of Big Data And Machine Learning", *NYU Journal of Intellectual Property and Entertainment Law*, 7(2): 78

Kaplan, Jerry. 2016. *Humans Need Not Apply, A Guide to Wealth and Work in the Age of Artificial Intelligence*, The USA: Yale University Press.

Kaminski, Margo. 2019. "The Right to Explanation, Explained"., *Berkeley Technology Law Journal* 34(1), 189. https://scholar.law.colorado.edu/articles/1227

Kaminski, Margo. 2019. "Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability" *Southern California Law Review* 92(6):1529–1616.

Kaminski, Margo and Gianclaudio Malgieri. 2020. "Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR", Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 68–79. https://doi.org/10.1145/3351095.3372875.

Kantardzic, Mehmed. 2015. *Data Mining: Concepts, Models, Methods, and Algorithms*, (2nd ed.) Hoboken, New Jersey: John Wiley.

Kasenberg , Daniel, Thomas Arnold and Matthias Scheutz. 2018. "Norms, Rewards, and the Intentional Stance: Comparing Machine Learning Approaches to Ethical Training" , Proceedings of the Thirty- Second AAAI Conference on Artificial Intelligence <https://hrilab.tufts.edu/ publications/kasenbergetal18aies.pdf>

Katz, Daniel Martin. 2012. "Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry" *Emory Law Journal*, Vol. 62, 2013. https://ssrn.com/abstract=2187752

Kauffman, Stuart A. 1971. "Articulation of parts explanations in biology" in R. C. Buck and R. S. Cohen (Eds.), *Boston Studies in the philosophy of science,* 8: 257–272.

Kelsen, Hans. 1941. "The Law as a Special Social Technique" *University of Chicago Law Review*: Vol. 9: Iss. 1, Article 5.

Kennedy, Rónán. 2020. "The Rule of Law and Algorithmic Governance." in Woodrow Barfield (ed.) *The Cambridge Handbook of the Law of Algorithms*, Cambridge: Cambridge University Press, 209–326.

Keplinger, Michael S. 1977. "Computer Intellectual Property Claims: Computer Software and Data Base Protection", *Washington University Law Quarterly*, 461.

Kerr, Ian and Jessica Earle. 2013. "Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy" 66 *Stanford Law Review Online*, 65-72.

Kessler, Friedrich, Edith Fine, 1964. "Culpa in Contrahendo, Bargaining in Good Faith, and Freedom of Contract: A Comparative Study." *Harvard Law Review* 77, no. 3: 401–49. https://doi.org/10.2307/1339028.

Kitchin, Rob. 2014. "Big Data, New Epistemologies and Paradigm Shifts", *Big Data & Society*1, no. 1 (1 April 2014): 205395171452848. https://doi.org/10.1177/2053951714528481.

Kleinberg J, Raghavan M (2018) How Do Classifiers Induce Agents To Invest Effort Strategically? arXiv: 1807.05307v5

Kleinberg, Jon, and Manish Raghavan. 2020. "How Do Classifiers Induce Agents to Invest Effort Strategically?", *ACM Transactions on Economics and Computation* 8, no. 4: 1–23. https://doi.org/10.1145/3417742.

Koivisto, Ida. 2016."The Anatomy of Transparency: The Concept and its Multifarious Implications", EUI Working Papers.

Koivisto, Ida. 2021. "Transparency in the Digital Environment" *Critical Analysis of Law* 8(1):1-8.

Kooiman, Jan (ed.). 1993. *Modern Governance: New Government-Society Interactions*. London; Newbury Park, Calif: Sage.

Koops, Bert-Jaap. 2007. "Criteria for Normative Technology: An essay on the acceptability of 'code as law in light of democratic and constitutional values", Tilburg University Legal Studies Working Paper No. 007/2007.

Koops, Bert-Jaap. 2013. "On Decision Transparency, or How to Enhance Data Protection after the Computational Turn" in Mireille Hildebrandt and K. de Vries (eds), *Privacy, Due Process and the Computational Turn* (Abingdon, Oxon, [England]; New York Routledge, 143-167. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/.

Korff, Douwe, 2010. "New Challenges to Data Protection Study - Working Paper No. 2: Data Protection Laws in the EU: The Difficulties in Meeting the Challenges Posed by Global Social and Technical Developments" European Commission DG Justice, Freedom and Security Report. http://dx.doi.org/10.2139/ssrn.1638949

Krenn, Karoline. 2017. “Introduction: Markets and Classifications - Constructing Market Orders in the Digital Age. An Introduction.” *Historical Social Research / Historische Sozialforschung* 42, no. 1 (159): 7–22. http://www.jstor.org/stable/44176022.

Kroll, Joshua, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. “Accountable Algorithms.” *University of Pennsylvania Law Review* 165(3): 633-705.

Kroll, Joshua A. 2018. "The Fallacy of Inscrutability", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2133: 20180084. https://doi.org/10.1098/rsta.2018.0084.

Kroll, Joshua A. 2020. "Accountability in Computer Systems" in Markus D. Dubber, Frank Pasquale, Sunit Das (eds), *The Oxford Handbook Of Ethics Of AI*, New York, NY: Oxford University Press, 180-196.

Krenn, Karoline. 2017. ‘Markets and Classifications – Constructing Market Orders in the Digital Age: An Introduction’ (2017) 42(1) Historical Social Research *42*(1), 7-22.

Kudyba, Stephan (ed). 2014. *Big Data, Mining, and Analytics: Components of Strategic Decision Making*. Boca Raton: Taylor & Francis.

Laat, Paul B. de. 2022. "Algorithmic Decision-Making Employing Profiling: Will Trade Secrecy Protection Render the Right to Explanation Toothless?" *Ethics and Information Technology* 24(2). https://doi.org/10.1007/s10676-022-09642-1.

La Diega, Guido Noto. 2018. "Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information" *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 9(3).

Latour, Bruno. 1993. *The Pasteurization of France* (Translated by Alan Sheridan, and John Law), Cambridge, Mass.: Harvard Univ. Press, 1993. (Originally published in 1984 as *Les microbes: guerre et paix suivi de imiductions*, A. M. Metailie, Paris)

Latzer, Michael and Noemi Festic. 2010. "A guideline for understanding and measuring algorithmic governance in everyday life" *Internet Policy Review* 8, no. 2 https://doi.org/10.14763/2019.2.1415.

Latzer, Michael, Katharina Hollnbuchner, Natascha Just, Florian Saurwein. 2016. "The Economics of Algorithmic Selection on the Internet" in Bauer, Johannes, and Michael Latzer (eds.) *Handbook on the Economics of the Internet*, Cheltenham, UK; Northampton, MA: Edward Elgar Publishing.

Le Bui, Matthew and Safiya Umoja Noble. 2020. "We’re Missing a Moral Framework of Justice in Artificial Intelligence: On the Limits, Failings, and Ethics of Fairness" in Markus D. Dubber, Frank Pasquale, Sunit Das (eds), *The Oxford Handbook Of Ethics Of AI*, New York, NY: Oxford University Press.

Leerssen, Paddy. 2022. "Algorithm Centrism in the DSA’s Regulation of Recommender Systems", Verfassungsblog, 22.03.2022, https://verfassungsblog.de/roa-algorithm-centrism-in-the-dsa/

Levy, Karen E.C. 2013. "Relational Big Data" 66 *Stanford Law Review Online* 73 (3).

Lemley, Mark A. 2004. "Ex Ante Versus Ex Post Justifications for Intellectual Property’", *University Chicago Law Review,* 2004, 71(1):129.

Lemley, Mark A. 2008. "The Surprising Virtues of Treating Trade Secrets As IP Rights" Stanford Law Review 61(2): 311-351.

Leenes, Ronald. 1999. *Hercules of Karneades: Hard cases in recht en rechtsinformatica,* Enschede: Twente University Press.

Leenes, Ronald. 2008. "Reply: Addressing the Obscurity of Data Clouds" in Mireille Hildebrandt and Serge Gutwirth (eds), *Profiling the European Citizen: Cross-disciplinary Perspectives,* Dordrecht: Springer, 341-351.

Leenes, Ronald. 2011. "Framing Techno-Regulation: An Exploration of State and Non-State Regulation by Technology". *Legisprudence* 5, no. 2): 143–69. https://doi.org/10.5235/175214611797885675.

Leese, Matthias. 2014. "The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union", *Security Dialogue* 45, no. 5.

Lehr, David and Paul Ohm. 2017. "Playing with the Data: What Legal Scholars Should Learn about Machine Learning" 51 *UC Davis Law Review,* 653.

Leibniz, Gottfried Wilhelm. 1979 [1685]. *Selections*, Philip P. Wiener (ed.), New York: Charles Scribner's Sons.

Leith, Philip. 2010. "The rise and fall of the legal expert system" vol.1 *European Journal of Law and Technology.*

Leistner, Matthias. 2017. "Big Data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform" in Sebastian Lohsse, Reiner Schulze, and Dirk Staudenmayer (eds). *Trading Data in the Digital Economy: Legal Concepts and Tools*. 1st edition. Baden-Baden, Germany: [Oxford, England]: Nomos ; Hart Publishing.

Lepri, Bruno, Jacopo Staiano, David Sangokoya, Emmanuel Letouzé and Nuria Oliver. 2016. "The tyranny of data? The bright and dark sides of data-driven decision-making for social good" arX- iv:1612.00323.

Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace,* New York: Basic Books.

Lettieri, Nicola. 2021. "Law in the turing's cathedral: Notes on the algorithmic turn of the legal universe." in Woodrow Barfield (ed.) *The Cambridge Handbook of the Law of Algorithms*, United Kingdom; New York, NY: Cambridge University Press, 691-721.

Li, Jundong, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P. Trevino, Jiliang Tang, and Huan Liu. 2018. "Feature Selection: A Data Perspective", *ACM Computing Surveys* 50(6): 1–45. https://doi.org/10.1145/3136625.

Li, Phoebe. 2021. "Intellectual property for humanity: A manifesto" in Daniel Gervais (ed.) *The Future of Intellectual Property*, Cheltenham, UK; Northampton, Massachusetts: Edward Elgar Publishing Limited, 9-37.

Lipton, Zachary C. 2016. "The Mythos of Model Interpretability" arXiv:1606.03490v3 https://doi.org/10.48550/arXiv.1606.03490

Lipton, Peter. 2004. *Inference to the Best Explanation,* London; New York: Routledge/Taylor and Francis Group.

Loi, Michele, Andrea Ferrario and Eleonora Viganò. 2021. "Transparency as Design Publicity: Explaining and Justifying Inscrutable Algorithms" *Ethics and Information Technology* 23: 253–263.

Lohsse, Sebastian, Reiner Schulze and Dirk Staudenmayer. 2017. "Trading Data in the Digital Economy: Legal Concepts and Tools" in Sebastian Lohsse, Reiner Schulze, Dirk Staudenmayer (eds) *Trading Data in the Digital Economy: Legal Concepts and Tools,* Baden-Baden, Germany: [Oxford, England]: Nomos; Hart Publishing, 2017), 13-24.

Lynskey, Orla. 2015. *The Foundations of EU Data Protection Law*, Oxford, United Kingdom: Oxford University Press.

Mackor, Anne Ruth. 2011. 'Explanatory Non-Normative Legal Doctrine. Taking the Distinction between Theoretical and Practical Reason Seriously', in Mark van Hoecke, (ed.) *Methodologies of Legal Research: Which Kind of Method for What Kind of Discipline?* Oxford; Portland, Or: Hart, 45-70.

Macmillan, Rory. 2018. "Big Data, Machine Learning, Consumer Protection and Privacy", Geneva, Switzerland: International Telecommunication Union (ITU) Security, Infrastructure and Trust Working Group.

Maggiolino, Mariateresa. 2019. "EU Trade Secrets Law and Algorithmic Transparency" Bocconi Legal Studies Research Paper No. 3363178, 9

Malgieri, Gianclaudio. 2016. "Trade Secrets v Personal Data: A possible solution for balancing rights" *International Data Privacy Law,6*(2):102–116.

Malgieri, Gianclaudio. 2016. "Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data" *PinG Privacy in Germany*, no. 4 :5. https://doi.org/10.37307/j.2196-9817.2016.04.05.

Malgieri, Gianclaudio, and Giovanni Comandé. 2017."Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation". *International Data Privacy Law* 7(4): 243–65. https://doi.org/10.1093/idpl/ipx019.

Malgieri, Gianclaudio. 2019. "Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations". *Computer Law & Security Review* 35, no. 5: 105327. https://doi.org/10.1016/j.clsr.2019.05.002.

Mantelero, Alessandro. 2018. "AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment" 34(4) *Computer Law & Security Review*,754-772.

Mantelero, Alessandro. 2019. "Artificial Intelligence and Data Protection: Challenges and Possible Remedies", The Council of Europe's Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. https://rm.coe.int/arti!cial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6.

Markou, Christopher and Simon Deakin. 2020. "Ex Machina Lex: Exploring the Limits of Legal Computability" in Simon Deakin and Christopher Markou (eds), *Is Law Computable: Critical Perspectives on Law and Artificial Intelligence*, Oxford: Hart Publishing, 31-66.

Maskus, Keith E. 2012. *Private Rights and Public Problems: the Global Economics of Intellectual Property in the 21st Century,* Washington, DC: Peterson Institute for International Economics.

Massé, Estelle and Laureline Lemoine. 2019. "One Year Under the GDPR", Access Now Publication. Available from URL: https://www.accessnow. org/cms/assets/uploads/2019/06/One-Year-Under-GDPR.pdf

Mattioli, Michael. 2014. "Disclosing Big Data", *Minnesota Law Review* 99(2): 535-583.

Matus, Kira J.M. and Michael Veale. 2022. "Certification Systems for Machine Learning: Lessons from Sustainability", *Regulation & Governance* 16(1): 177–196. https://doi.org/10.1111/rego.12417.

Mayer-Schönberger, Viktor and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt Publishing Company.

McMahon Aisling. 2019. "Gene Patents and the Marginalisation of Ethical Issues", *European Intellectual Property Review* 41(10):608–620.

McIntyre, T. J. and Scott, Colin David, 2008. "Internet Filtering: Rhetoric, Legitimacy, Accountability and Responsibility" in Brownsword, Roger and Yeung, Karen, (eds.) *Regulating Technologies Legal: Futures, Regulatory Frames and Technological Fixes*. Oxford, UK: Hart Publishing, 109-125. https://ssrn.com/abstract=1103030

Mendoza, Isak and Lee Bygrave. 2017. "The Right Not to Be Subject to Automated Decisions Based on Profiling" in Tatiani Synodinou, Philippe Jougleux, Christiana Markou, Thalia Prastitou (eds.), *EU Internet Law: Regulation and Enforcement,* New York, NY: Springer Berlin Heidelberg, 77-98.

Micova, Sally Broughton, 2022. "DMA:Transparency Requirements in Relation to Advertising", Issue Paper, November 2022 CERRE project entitled 'Effective and Proportionate Implementation of the DMA' https://cerre.eu/wp-content/uploads/2022/11/DMA_TransparencyRequirementsinAdvertising.pdf

Millar, Jason and Ian Kerr, "Delegation, Relinquishment and Responsibility: The Prospect of Expert Robots" in Ryan Calo, Michael Froomkin and Ian Kerr (eds), *Robot Law*, Cheltenham, UK:Edward Elgar, 102–218.

Miller, Tim. 2019."Explanation in Artificial Intelligence: Insights from the Social Sciences", *Artificial Intelligence* 267:1–38. https://doi.org/10.1016/j.artint.2018.07.007.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji and Timnit Gebru. 2019. "Model Cards for Model Reporting." *Proceedings of the Conference on Fairness, Accountability, and Transparency*.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. "The Ethics of Algorithms: Mapping the Debate" *Big Data & Society* 3, no. 2.

Moses, Lyria Bennett. 2013. "How to Think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target", 5 *Law, Innovation and Technology*, 1-20.

Moses, Lyria Bennett and Janet Chan. 2014. 'Using Big Data for Legal and Law Enforcement Decisions' 37 *New South Wales Law Journal*, 643-678.

Muller, Scott J. 2007. *Asymmetry: The Foundation of Information*, Berlin: Springer.

Murray, Andrew and Colin Scott. 2002. "Controlling the New Media: Hybrid Responses to New Forms of Power" *The Modern Law Review* 65(4): 491-516.

Murray, Andrew D. 2008. "Conceptualising the post-regulatory (cyber)state." in Brownsword, Roger and Yeung, Karen, (eds.) *Regulating Technologies Legal: Futures, Regulatory Frames and Technological Fixes*. Oxford, UK: Hart Publishing, 287-316.

Neuwirth, Rostam J. 2018. *Law in the Time of Oxymora: A Synesthesia of Language, Logic and Law*. London; New York, NY: Routledge Taylor and Francis Group.

O'Connor, Sean M. 2008. "Enabling Research or Unfair Competition? De Jure and De Facto Research Use Exceptions in Major Technology Countries" in Toshiko Takenaka (ed), *Patent Law & Theory: A Handbook Of Contemporary Research,* Cheltenham, UK ; Northampton, MA: Edward Elgar Publishing, 519-567.

Oksanen, Kenneth, Perttu Virtanen, Eljas Soisalon-Soininen, Jukka Kemppinen. 2011. "Arguments in Considering the Similarity of Algorithms in Patentin" *SCRIPTed,* 8(2):138-153.

Page, Scott E. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton, NJ: Princeton University Press.

Paliwala, Abdul. 2016. "Rediscovering artificial intelligence and law: an inadequate jurisprudence?" 30 *International Review of Law, Computers & Technology* 107-114.

Palmiotto, Francesca. 2021. "The Black Box on Trial: The Impact of Algorithmic Opacity on Fair Trial Rights" in Martin Ebers and Marta Cantero Gamito (eds), *Algorithmic Governance and Governance of Algorithms: Legal and Ethical Challenges. Data Science, Machine Intelligence, and Law,* volume 1. Cham: Springer, 49-70.

Pasquale, Frank. 2011. "Restoring Transparency to Automated Authority" *Journal on Telecommunications and High Technology Law*, 9: 235-254. https://ssrn.com/abstract=1762766

Pasquale, Frank. 2011. "The troubling consequences of trade secret protection of search engine rankings" in Rochelle C. Dreyfuss and Katherine J. Strandburg (eds) *The Law and Theory of Trade Secrecy*, Cheltenham, UK; Northampton, MA: Edward Elgar, 381-405.

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

Pasquinelli, Matteo. 2019. "How a Machine Learns and Fails" *spheres: Journal for Digital Cultures*. *Spectres of AI* (2019), No. 5:1–17. DOI: https://doi.org/10.25969/mediarep/13490.

Paßmann, Johannes and Asher Boersma. 2017. "Unknowing Algorithms: On Transparency of Unopenable Black Boxes" in Mirko Tobias Schäfer and Karin van Es (eds.), *The Datafied Society: Studying Culture through Data*. Amsterdam: Amsterdam University Press, 139–146.

Pavlakos, George and Veronica Rodriguez-Blanco (eds). 2015. *Reasons and Intentions in Law and Practical Agency*. New York: Cambridge University Press.

Pearl, Judea and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.

Peczenik, Aleksander. 2005. "Scientia Juris. Legal Doctrine as Knowledge of Law and as a Source of Law", in Enrico Pattaro (ed), *A Treatise of Legal Philosophy and General Jurisprudence*, Dordrecht: Springer.

Peeters, Rik. 2020. "The Agency of Algorithms: Understanding Human-Algorithm Interaction in Administrative Decision-Making" *Information Polity*, 25(4): 507–522.

Peukert, Alexander. 2015."The Fundamental Right to (intellectual) property" in Christophe Geiger (ed.), *Research Handbook on Human Rights and Intellectual Property*, Cheltenham, UK: Edward Elgar 132-148, https://ssrn.com/abstract=2324132 or http://dx.doi.org/10.2139/ssrn.2324132

Polanyi, Karl. [1944] 2001. *The great transformation: The political and economic origins of our time*. Boston: Beacon Press.

Purtova, Nadezhda. 2018."The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law", *Law, Innovation and Technology* 10, no. 1: 40–81. https://doi.org/10.1080/17579961.2018.1452176.

Radbruch, Gustav. 1950. "Legal Philosophy" in *The Legal Philosophies of Lask, Radbruch and Dabin* (Edwin W. Patterson (ed.), Kurt Wilk trans.) Cambridge, Massachusetts :Harvard University Press.

Rader, Emilee, Kelley Cotter, and Janghee Cho. 2018. "Explanations as Mechanisms for Supporting Algorithmic Transparency" in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3173574.3173677.

Radoń, Barbara Anna. "*Trade Secrets Protection for 'Big Data': Personal Data as Trade Secrets in the European Union"* (Master thesis, MIPLC 2015/16)

Raue, Benjamin. 2018. "Free Flow of Data? The Friction Between the Commission's European Data Economy Initiative and the Proposed Directive on Copyright in the Digital Single Market" *IIC - International Review of Intellectual Property and Competition Law* 49 (4): 379–83.

Raz, Joseph. 2009. *The Authority of Law: Essays on Law and Morality*, 2nd ed. Oxford New York: Oxford University Press.

Ribeiro, Marco Tulio, Sameer Singh and Carlos Guestrin, 2016."Why should I trust you?: Explaining the predictions of any classifier" Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 13–17 August 2016.

Ribeiro, Gonçalo de Almeida. 2019. *The Decline of Private Law: A Philosophical History of Liberal Legalism*. Law and Practical Reason. Oxford, UK; Chicago, Illinois: Hart Publishing.

Ricketson, Sam. 1992. "The 1992 Horace S. Manges Lecture -People or Machines: The Berne Convention and the Changing Concept of Authorship", 16 *The Columbia Journal of Law & the Arts* (1991-1992).

Roberts, John. 2009. "No One is Perfect: The Limits of Transparency and an Ethic for 'Intelligent' Accountability" *Accounting, Organizations and Society*, 34(8): 957–970. https://doi.org/10.1016/j.aos.2009.04.005.

Robinson, Timothy D. 2017. "A Normative Evaluation of Algorithmic Law", 23 *Auckland University Law Review,* 293-323.

Robnik-Šikonja, Marko and Igor Kononenko. 2008. "Explaining Classifications for Individual Instances" 20 IEEE Transactions on Knowledge and Data Engineering

Rosenberg, Daniel. 2013. "Data before the Fact" in Lisa Gitelman (ed.) *"Raw Data" Is an Oxymoron*, Cambridge, Massachusetts: MIT Press. 15-40.

Roosendaal*,* Arnold. 2013. *Digital personae and profiles in law: Protecting individuals' rights in online contexts,* Oirschot:Wolf Legal Publishers

Rouse, Joseph. 2007. "Social Practices and Normativity", *Philosophy of the Social Sciences* 37(1): 46–56. https://doi.org/10.1177/0048393106296542

Rouvroy, A. Antoinette. 2013. "The end(s) of critique: Data behaviourism versus due process" in Mireille Hildebrandt and K de Vries (eds), *Privacy, Due Process and the Computational Turn*, Abingdon, Oxon, [England]; New York Routledge*,* 143-167.

Ruiter, Dick W. P. 1993. *Institutional Legal Facts: Legal Powers and Their Effects*. Law and Philosophy Library, v. 18. Dordrecht ; Boston : Norwell, MA, U.S.A: Kluwer Academic Publishers.

Ruger, Theodore W., Pauline T. Kim, Andrew D. Martin, and Kevin M. Quinn. 2004. "The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking" 104 *Columbia Law Review* no.4, 1150.

Sadowski, Jathan. 2020. *Too Smart: How Digital Capitalism Is Extracting Data, Controlling Our Lives, and Taking over the World*. Cambridge, Massachusetts: MIT Press.

Samaha, Adam. 2008. "Judicial Transparency in an Age of Prediction", University of Chicago Public Law & Legal Theory Working Paper No. 216.

Samuelson, Pamela. 1986. "Allocating Ownership Rights in Computer-Generated Works" 47 *University of Pittsburgh Law Review*, 1185-1228.

Samuelson, Pamela. 2007. "Why Copyright Law Excludes Systems and Processes From Its Scope of Protection" *Texas Law Review,* 85(7): 1921-1977.

Samuelson, Pamela. 2017. "Functionality and Expression in Computer Programs: Refining the Tests for Software Copyright Infringement" *Berkeley Technology Law Journal* 31 (3):1215-1300.

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms" paper presented at *Data and Discrimination: Converting Critical Concerns into Productive Inquiry.* a preconference at the 64th Annual Meeting of the International Communication Association. May 22, 2014; Seattle, WA, USA.

Sappa, Cristiana, '2018), "What Does Trade Secrecy Have to Do with the Interconnection-based paradigm of the Internet of Things?" *European Intellectual Property Review*, 40(8):518-523.

Schartum, Dag Wiese. 2016. "Making Privacy by Design Operative", *International Journal of Law and Information Technology* 24(2):151–175. https://doi.org/10.1093/ijlit/eaw002

Schönberger, Daniel. 2018. "Deep Copyright: Up - And Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML)" in Jacques De Werra (ed.), *Droit d'auteur 4.0 / Copyright 4.0,* Geneva / Zurich, Schulthess Editions Romandes, 145-173. https://ssrn.com/abstract=3098315

Selbst, Andrew D. 2017. "A Mild Defense of Our New Machine Overlords" 70 *Vanderbilt Law Review En Banc*, 87-105. https://ssrn.com/abstract=2941078

Selbst, Andrew D. and Solon Barocas. 2018. "The Intuitive Appeal of Explainable Machines", *Fordham Law Review* 87, 1085–1139.

Selbst, Andrew D., Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. "Fairness and abstraction in sociotechnical systems" in Proceedings of the *Conference on Fairness, Accountability, and Transparency*, 59–68. ACM.

Seltzer, Wendy. 2013. "Software Patents and/or Software Development", *Brooklyn Law Review* 78(3): 929-987.

Shakespeare, William. 2003. *Hamlet, Prince of Denmark*, Philip Edwards (ed) Cambridge, U.K.; New York: Cambridge University Press.

Simon, Brenda and Ted Sichelman. 2017. "Data-Generating Patents" *Northwestern University Law Review,* 111: 377-437,: https://ssrn.com/abstract=2753547

Simpson, Thomas W. 2014. "Evaluating Google As An Epistemic Tool" in Harry Halpin and Alexandre Monnin (eds) *Philosophical Engineering Toward a Philosophy of the Web,* Chichester, West Sussex, UK*:* Wiley Blackwell, 97-116.

Singh, Jatinder, Ian Walden, Jon Crowcroft, and Jean Bacon. 2016. "Responsibility & Machine Learning: Part of a Process" *SSRN Electronic Journal*, 2016. https://doi.org/10.2139/ssrn.2860048.

Skelton, Sebastian Klovig. 2021. "Europe's Proposed AI Regulation Falls Short on Protecting Rights.". Computer Weekly.com. https://www.computerweekly.com/feature/Europes-proposed-AI-regulation-falls-short-on-protecting-rights

Smuha, Nathalie, Emma Ahmed-Rengers, Adam Harkens, Wenlong Li,d James MacLaren,eRiccardo Pisellif and Karen Yeung. 2021. "How the EU can achieve legally trustworthy AI : a response to the European Commission's proposal for an Artificial Intelligence Act", *Artificial Intelligence - Law, Policy, & Ethics eJournal*. https://ssrn.com/abstract=3899991

Sobel, Benjamin L. W. 2017. "Artificial Intelligence's Fair Use Crisis". *The Columbia Journal of Law & The Arts* 41 (1):45-97. https://doi.org/10.7916/jla.v41i1.2036.

Sousa e Silva, N. 2014."What Exactly Is a Trade Secret under the Proposed Directive?' *Journal of Intellectual Property Law & Practice* 9(11) (1 November 2014): 923–32. https://doi.org/10.1093/jiplp/jpu179.

Spielkamp, Matthias (ed.). 2019. *Automating Society Taking Stock of Automated Decision-Making in the EU*, AlgorithmWatch, Berlin. https://algorithmwatch.org/en/wpcontent/uploads/2019/02/Automating_Society_Report_2019.pdf

Stalder, Felix. 2018. "From inter-subjectivity to multi-subjectivity. Knowledge claims and the digital condition" in Emre Bayamlıoğlu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt, (eds),*BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen*, Amsterdam: Amsterdam University Press, 98-101. https://doi.org/10.1515/9789048550180-018

Stamatoudi, Irini A., "Text and Data Mining" in Irini A. Stamatoudi (ed.), *New Developments in EU and International Copyright Law,* Leiden, Netherlands: Kluwer Law International, 251-282.

Sterckx, Sigrid, and Julian Cockbain. 2015. *Exclusions from Patentability How Far Has the European Patent Office Eroded Boundaries?* Cambridge: Cambridge University Press.

Stranieri, Andrew and John Zeleznikow. 2005. *Knowledge discovery from legal databases*, Dordrecht: Springer.

Svensson, Jörgen S. 2003. "Legal Expert Systems in Social Administration: From Fearing Computers to Fearing Accountants", *Information Polity* 7, no. 2,3 (20 March 2003): 143-154.

Surblytė, Gintarė. 2016. "Enhancing TRIPS: Trade Secrets and Reverse Engineering" in Hanns Ullrich, Reto M. Hilty, Matthias Lamping, and Josef Drexl (eds), *TRIPS plus 20: From Trade Rules to Market Principles*. Berlin, Heidelberg: Springer, Berlin, Heidelberg, 725-760.

Susskind, Richard E. and Daniel Susskind. 2015. *The Future of the Professions: How Technology Will Transform the Work of Human Experts*. First edition. Oxford, United Kingdom: Oxford University Press.

Synodinou, Tatiana-Eleni. 2020. "Who Is a Lawful User in European Copyright Law? From a Variable Geometry to a Taxonomy of Lawful Use" in Tatianē-Elenē Synodinou, Philippe Jougleux, Christiana Markou, and Thalia Prastitou (eds.) *EU Internet Law in the Digital Era: Regulation and Enforcement*. Cham, Switzerland: Springer, 27-60.

Tamanaha, Brian Z. 2004. *On the Rule of Law: History, Politics, Theory*. Cambridge; New York: Cambridge University Press.

Tamanaha, Brian Z. 2017. *A Realistic Theory of Law*. Cambridge, United Kingdom: Cambridge University Press.

Taylor, Linnet. 2016. *Group Privacy: New Challenges of Data Technologies*. New York, NY: Springer Berlin Heidelberg,

Taylor, Linnet, Luciano Floridi and Bart van der Sloot (eds). 2017. *Group Privacy: New Challenges of Data Technologies,* Dordrecht: Springer.

Teunissen, Peter. 2018."The Balance Puzzle: the ECJ's Method of Proportionality Review for Copyright Injunctions" *European Intellectual Property Review*. 40(9):579-593.

Thagard, Paul. 1988, *Computational Philosophy of Science,* Cambridge, Massachusetts: MIT Press.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Tomkowicz, Robert. 2012. *Intellectual Property Overlaps: Theory, Strategies and Solutions*. New York: Routledge.

Tramer, Florian, Fan Zhang, Ari Juels, Michael K. Reiter, nd Thomas Ristenpart.2016. “Stealing Machine Learning Models via Prediction APIs”, 25th USENIX Security Symposium, USENIX Association, 2016) 601–618. https://floriantramer.com/docs/papers/sec16stealing.pdf

Triaille, Jean-Paul, Jérôme de Meeûs D’Argenteuil and Amélie de Francquen. 2014. *Study on the Legal Framework of Text and Data Mining (TDM),* European Commission, Directorate-General for the Internal Market and Services, Publications Office. https://data.europa.eu/doi/10.2780/1475

Ullrich, Hanns, Reto M. Hilty, Matthias Lamping, and Josef Drexl, eds. 2016. *TRIPS plus 20: From Trade Rules to Market Principles*. Berlin, Heidelberg: Springer, https://doi.org/10.1007/978-3-662-48107-3.

Vaesen, Krist. 2006. "How Norms in Technology Ought to Be Interpreted", *Techné: Research in Philosophy and Technology* 10, no.1: 95–108. https://doi.org/10.5840/techne200610144.

Vale, Sebastião Barros and Gabriela Zanfir-Fortuna, 2022. "Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities" Future of Privacy Forum report, 19.

van den Berg, Bibi and Ronald Leenes. 2013. "Abort, retry, fail: scoping techno-regulation and other techno-effects" in Mireille Hildebrandt and Jaenne Gakeer (eds), *Human Law and Computer Law: Comparative Perspectives,* New York: Springer.

van Hoecke, Mark. 2011. ‘Legal Doctrine: Which Method(s) for What Kind of Discipline?’ in Mark van Hoecke, ed. *Methodologies of Legal Research: Which Kind of Method for What Kind of Discipline?* Oxford; Portland, Or: Hart 1-19.

van Gompel, Stef. 2014."Creativity, autonomy and personal touch. A critical appraisal of the CJEU’s originality test for copyright" in Mireille van Eechoud (ed.) *The Work of Authorship*, Amsterdam: Amsterdam University Press, 95-144. https://doi.org/10.1515/9789048523009-004

van Otterlo, Martijn. 2018. "Gate-keeping Algorithms with Human Ethical Bias", arX- iv:1801.01705v1. https://arxiv.org/abs/1801.01705.

Varošanec, Ida. 2022. "On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI", *International Review of Law, Computers & Technology*, 36:2, 95-117, DOI: 10.1080/13600869.2022.2060471.

Veale, Michael and Frederik J. Zuiderveen Borgesius. 2021. "Demystifying the Draft EU Artificial Intelligence Act – Analysing the Good, the bad, and the Unclear Elements of the Proposed Approach." *Computer Law Review International* 22 (4): 97–112.

Veale, Michael, Reuben Binns, and Lilian Edwards. 2018. "Algorithms That Remember: Model Inversion Attacks and Data Protection Law'. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 213: 20180083. https://doi.org/10.1098/rsta.2018.0083.

Vedder, Anton, and Laurens Naudts. 2017. "Accountability for the Use of Algorithms in a Big Data Environment", *International Review of Law, Computers & Technology* 31, no. 2: 206–224.

Vedder, Anton. 2018. "Why data protection and transparency are not enough when facing social problems of machine learning in a big data context" in Emre Bayamlıoğlu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt, (eds), *BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen*, Amsterdam: Amsterdam University Press, 42-45. DOI 10.5117/9789463722124/CH6.

Verbeek, Paul. 2005. What Things Do – Philosophical Reflections on Technology, Agency, and Design (Robert P Crease, trans.) University Park, Pa: Pennsylvania State University Press.

Vries, Marc J. de, Sven Ove Hansson, and Anthonie Meijers, (eds). 2013. *Norms in Technology*. Philosophy of Engineering and Technology, v. 9. Dordrecht; New York: Springer.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation", *International Data Privacy Law* 7(2): 76–99. https://doi.org/10.1093/idpl/ipx005.

Wachter, Sandra, Brent Mittelstadt, and Chris Russell. 2018. "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR" *Harvard Journal of Law & Technology* 31(2), 841–887.

Wachter, Sandra and Brent Mittelstadt, 2019 "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI" *Columbia Business Law Review*, *2019*(2), 494–620. https://doi.org/10.7916/cblr.v2019i2.3424

Waldman, Ari Ezra. 2021. *Industry Unbound: The inside Story of Privacy, Data, and Corporate Power*, Cambridge, United Kingdom ; New York, NY, USA:Cambridge University Press.

Waldron, Jeremy. 1989. "Rights in Conflict", *Ethics,* 99(3)*:* 503–519.

Waldron, Jeremy. 1989. "The Rule of Law in Contemporary Liberal Theory", *Ratio Juris* 2(1): 79–96.

Waldron, Jeremy. 2010. "The Rule of Law and the Importance of Procedure", NYU School of Law, Public Law Research Paper No. 10-73.

Weller, Adrian. 2017. "Challenges for Transparency", *ICML Workshop on Human Interpretability in Machine Learning*, Sydney, NSW, Australia. https://arxiv.org/abs/1708.01870.

Wiebe, Andreas. 2017. "Protection of Industrial Data – a New Property Right for the Digital Economy?' *Journal of Intellectual Property Law & Practice* 12(1): 62–71. https://doi.org/10.1093/jiplp/jpw175.

Wimsatt, William.1972. "Complexity and organization" in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 20: 67–86, Dordrecht: D. Reidel

Wimsatt, William. 2007. *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*, Cambridge, Mass: Harvard University Press.

Winner, Langdon. 1977. *Of Autonomous Technology: Technics-Out-of-Control as a Theme in Political Thought,* Cambridge, Mass: MIT Press.

Wischmeyer, Thomas. 2020. "Artificial Intelligence and Transparency: Opening the Black Box" in Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence*. Cham, Switzerland: Springer, 75–101.

Yeung, Karen. 2008. "Towards an Understanding of Regulation by Design" in R Brownsword and K Yeung (eds), *Regulating Technologies Legal: Futures, Regulatory Frames and Technological Fixes*. Oxford, UK: Hart Publishing, 79-107.

Yeung, Karen. 2015. " Design for Regulation" in Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel, eds. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*. Dordrecht Heidelberg Heidelberg New York London: Springer, 447–472.

Yeung Karen. 2017. “Hypernudge: big data as a mode of regulation by design" *Inf Commun Soc* 20:118.

Yeung, Karen. 2018. "Algorithmic Regulation: A Critical Interrogation", *Regulation & Governance* 12(4), 505–523.

Yeung, Karen and Martin Lodge (eds.). 2019. *Algorithmic Regulation,* Oxford: Oxford University Press.

Zalnieriute, Monika, Lyria Bennett Moses, George Williams. 2019. “The Rule of law and Automation of Government Decision-Making.” *The Modern Law Review* 82 (3): 425–455

Zarsky, Tal. 2013. "Transparent Predictions", *University of Illinois Law Review*, 1503-1569.

Zarsky, Tal. 2017. "Incompatible: The GDPR in the Age of Big Data", *Seton Hall Law Review* 47, 995–1020.

Zech, Herbert. 2016. "A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data", *Journal of Intellectual Property Law & Practice* 11(6): 460–70. https://doi.org/10.1093/jiplp/jpw049.

Zerilli, John, Alistair Knott, James Maclaurin, and Colin Gavaghan. 2019. "Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?" *Philosophy and Technology, 32* (4), 661-683. https://doi.org/10.1007/s13347-018-0330-6

Zerzan, John. 2015. *Why hope?: the stand against civilization*, Port Townsend (Wash.): Feral House.

Zhiwei, Zeng, Xiuyi Fan, Chunyan Miao, Cyril Leung, Chin Jing Jih, and Ong Yew Soon. 2018. "Context-based and Explainable Decision Making with Argumentation" in Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1114–1122.